

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- BLURRY OR ILLEGIBLE TEXT
- SKEWED/SLATED IMAGES
- COLORED PHOTOS
- BLACK OR VERY DARK BLACK AND WHITE PHOTOS
- UNDECIPHERABLE GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

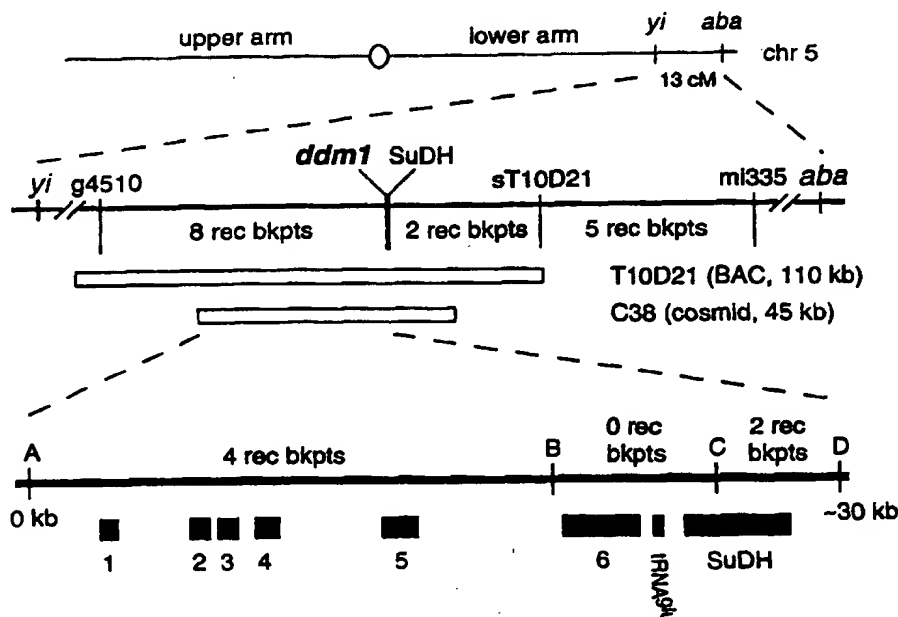
THIS PAGE BLANK (USPTO)



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/82, C07K 14/415, 16/16, A01H 5/00		(11) International Publication Number: WO 99/55891
A1		(43) International Publication Date: 4 November 1999 (04.11.99)
(21) International Application Number: PCT/US99/09268		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW. ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 29 April 1999 (29.04.99)		
(30) Priority Data: Not furnished 30 April 1998 (30.04.98) US 09/104,070 24 June 1998 (24.06.98) US		
(71) Applicant (for all designated States except US): WASHINGTON UNIVERSITY [US/US]; 600 South Euclid Avenue, St. Louis, MO 63110 (US).		
(72) Inventors; and (75) Inventors/Applicants (for US only): RICHARDS, Eric, J. [US/US]; 4446 Westminster Place, St. Louis, MO 63108 (US). JEDDELOH, Jeffrey, A. [US/US]; 6756 W. Lakeridge Drive, New Market, MD 21774 (US).		
(74) Agents: REED, Janet, E. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103 (US).		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PLANT GENE THAT REGULATES DNA METHYLATION



(57) Abstract

A novel gene, *DDM1*, and its encoded protein are provided. The gene was isolated from a region of *Arabidopsis thaliana* chromosome 5. *DDM1* appears to be part of the SWI2/SNF2 family of chromatin-remodeling proteins. Disruption of the gene results in DNA hypomethylation, among other phenotypes. The *DDM1* gene defines a novel member of the DNA methylation system. Methods of using *DDM1*, and transgenic organisms comprising *DDM1*, are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PLANT GENE THAT REGULATES DNA METHYLATION

Pursuant to 35 U.S.C. §202(c), it is acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the National Science Foundation, Grant
5 Nos. MCB9306266 and BIR9256779.

This application claims priority to U.S. Provisional Application Serial No. 60/_____, filed April 30, 1998, and to U.S. Application No. 09/104,070, filed June 24, 1998 the entireties of which are
10 incorporated by reference herein.

FIELD OF THE INVENTION

This invention relates to the field of plant molecular biology, genetic engineering and regulation of
15 gene expression. In particular, this invention provides a novel gene, *DDM1*, which plays an important role in the regulation of DNA methylation, and resultant regulation of gene expression, in plant genomic DNA.

20 BACKGROUND OF THE INVENTION

Various publications or patents are cited in this application to describe the state of the art to which the invention pertains. Each of these publications or patents is incorporated by reference herein.

25 Plant genomes contain substantial amounts of 5-methylcytosine. Up to 20-30% of the cytosines are methylated in the nuclear genome of many flowering plants. As in other organisms, methylation of cytosine

- 2 -

residues in plants occurs post-replicatively through the action of cytosine-DNA methyltransferases. Plant DNA methyltransferases have been characterized biochemically, and plant genes encoding these enzymes have been isolated
5 by virtue of their similarity to their mammalian counterparts.

Investigations of native plant genes and transgenic plants containing foreign genes have found a general correlation between transcriptional inactivity
10 and increased DNA methylation, consistent with evidence from mammalian systems. This evidence supports a role for cytosine methylation in maintaining transcriptional states.

The plant's need for developmental plasticity
15 and environmental interaction suggests that plants extensively employ epigenetic regulatory strategies. Such strategies rely on heritable, often reversible, changes in access to the underlying genetic information, but not alteration of the primary nucleotide sequence.
20 As one example, the alteration of DNA methylation is expected to perturb plant development significantly, provided that differential DNA methylation is an important component of epigenetic regulation in plants.

One paradigm linking DNA methylation and
25 developmental regulation comes from work on the mouse, where average genome cytosine methylation levels in embryonic lineages drop sharply in the early cleavages following fertilization, then rise again around the time of implantation. In plants, a similar pattern has been
30 observed in studies of DNA methylation content in pollen and post-embryonic tissue of varying age. Information from such studies indicates that there is a gradual rise in 5-methylcytosine levels in post-embryonic tissues

- 3 -

produced by meristems at positions further from the base of the plant (i.e., tissues of increasing age). Genetic studies of transposon systems in maize also demonstrate an age-dependent gradient of increasing epigenetic
5 modification, which is correlated with DNA methylation.

Both biochemical and genetic approaches have been taken to alter DNA methylation in eucaryotic organisms. Methylation inhibitor treatments have induced developmental abnormalities in many plant species.
10 Transgenic plants expressing antisense molecules specific for a native cytosine methyltransferase gene have been found to exhibit genomic hypomethylation, presumably due to the antisense interference with expression of the gene.

15 In another approach, mutants of *Arabidopsis thaliana* have been isolated, which show a decrease in DNA methylation (*ddm*) resulting in reduced nuclear 5-methylcytosine levels. The best characterized mutations define the *DDM1* gene. Homozygotes carrying recessive
20 *ddm1* alleles contain 30% of the wild-type levels of 5-methylcytosine. The *ddm1* mutations do not map to the two known cytosine-DNA methyltransferase genes of *A. thaliana*, nor do they affect DNA methyltransferase activity detectable in nuclear extracts (Kakutani et al.,
25 *Nuc. Acids Res.* 23: 130-137, 1995). In addition, *ddm1* mutations do not appear to affect the metabolism of the active methyl group donor, S-adenosylmethionine (Kakutani et al., 1995, *supra*).

For the foregoing reasons, the *DDM1* gene
30 product is likely to be a novel component of the DNA methylation system, or involved in determining the cellular context (e.g., chromatin structure, subnuclear localization) of the methylation reaction. Consequently,

- 4 -

it would be a clear advance in the art of plant molecular and cellular biology to identify and isolate the *DDM1* gene and/or its encoded protein. Such a gene and protein would find utility for the purpose of modifying the methylation status of a selected genome and thereby altering one or more regulatory features of gene expression from that genome.

SUMMARY OF THE INVENTION

A novel gene, *DDM1*, and its encoded protein are provided in accordance with the present invention. The gene has been identified as a novel element of the DNA methylation system.

In one aspect of the invention, an isolated nucleic acid molecule comprising a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, is provided. The gene occupies a segment of chromosome 5, lower arm, which is flanked on the centromeric side within 20 kilobases by a gene encoding a zinc-finger protein and on the telomeric side within 1 kilobase by a gene encoding a glutamic acid tRNA. Disruption of the gene is associated with DNA hypomethylation. The gene encodes a polypeptide of about 764 amino acids in length. The nucleotide sequence of the *DDM1* gene is set forth herein as SEQ ID NO:1 and its deduced amino acid sequence as SEQ ID NO:2. In SEQ ID NO:1, the regions of the gene that comprise coding sequence are indicated.

In another aspect of the invention, an isolated *DDM1* gene is provided, having a sequence selected from the group consisting of: (a) SEQ ID NO:1; (b) an allelic variant or natural mutant of SEQ ID NO:1; (c) a sequence hybridizing with part or all of SEQ ID NO:1 or its complement and encoding a polypeptide substantially the

- 5 -

same as part or all of a polypeptide encoded by SEQ ID NO:1; (d) a sequence encoding part or all of a polypeptide having amino acid SEQ ID NO:2; and (e) a sequence encoding part or all of a polypeptide contained
5 in the cosmid clone C38, designated ATCC Accession No. 207208.

According to another aspect of the invention, a polypeptide is provided, which is produced by expression of an isolated nucleic acid molecule comprising part or
10 all of an open reading frame of a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, the gene occupying a segment of chromosome 5, lower arm, flanked on the centromeric side within 20 kilobases by a gene encoding a zinc-finger protein and on the telomeric side
15 within 1 kilobase by a gene encoding a glutamic acid tRNA. This polypeptide preferably has the amino acid sequence of part or all of SEQ ID NO:2.

According to another aspect of the invention, an isolated protein encoded by an *Arabidopsis thaliana*
20 gene is provided, which is a member of an SWI2/SNF2 family of polypeptides. Loss of function of the protein is associated with DNA hypomethylation. The protein is encoded by a gene located on *A. thaliana* chromosome 5, lower arm, centromerically flanked within 20 kilobases by
25 a zinc finger-encoding gene and telomerically within one kilobase by a gene encoding a glutamic acid tRNA.

According to another aspect of the invention, a transgenic organism comprising the *DDM1* gene is provided. In one embodiment, the transgenic organism is a plant.

30 In other aspects of the invention, methods are provided for stabilizing fidelity of DNA methylation in an organism, which comprise transforming the organism with the *DDM1* gene. Methods are also provided for

- 6 -

reducing or eliminating gene silencing in a plant, or for inducing inbreeding depression in a plant, which comprise inhibiting or preventing expression of an endogenous *DDM1* gene of the plant.

5 These aspects of the invention, as well as other features and advantages of the invention, will be described in greater detail in the description and examples set forth below.

10 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Map-based isolation of the *A. thaliana* *DDM1* gene. A genetic map of the region of *A. thaliana* chromosome 5 containing the *DDM1* gene is shown at the top of the figure (see Example 1). The relative
15 sizes of the genetic intervals were determined by the number of recombination breakpoints (rec bkpts) scored in a panel of recombinant lines containing cross-overs between flanking markers *yi* and *aba*. The regions represented in genomic clones T10D21 and C38 are denoted
20 by the open boxes below the genetic map. The ~30 kb interval containing the *DDM1* gene, defined by the genetic markers A and D, is shown at the bottom of the figure. The number of recombination breakpoints scored between markers A - D and *ddm1-2* are indicated. The position of
25 predicted coding regions in the interval are numbered and shown below the physical map. BAC, bacterial artificial chromosome; SuDH, succinate dehydrogenase structural gene.

Figure 2. *DDM1* gene structure and
30 identification. **Fig. 2A:** The intron/exon structure of the *DDM1* gene. Protein-coding exons are shown as open boxes, with the start and stop codons indicated. Introns are depicted as thin lines. The position and nature of

- 7 -

four *ddm1* alleles are indicated above the exon/intron map. **Fig. 2B:** RT-PCR analysis of *ddm1-2* and wild-type *DDM1* transcripts. The approximate positions of oligonucleotide primers used in the analysis are shown below the map in Fig. 2A. Amplifications were done on either genomic templates (DNA), first-strand cDNA templates (+RT, plus reverse transcriptase), or mock-synthesized cDNA (-RT, minus reverse transcriptase). Amplified products were separated on a 3% agarose gel and visualized after ethidium bromide staining. Amplification from cDNA representing the properly spliced transcript resulted in a ~280 bp product. The nucleotide sequence of the ~220 bp product amplified from *ddm1-2* cDNA template indicated that the mutation leads to use of an alternate splice donor 56 bp upstream of the wild-type splice donor site.

Figure 3. The *A. thaliana DDM1* gene encodes a SWI2/SNF2-like protein. The deduced primary amino acid sequence of *DDM1* (At *DDM1*) is aligned with two other SWI2/SNF2-like protein sequences, *Mus musculus* lymphocyte specific helicase (Mm LSH; SEQ ID NO:4) and human SNF2h (Hs SNF2h; SEQ ID NO:5). Sequence identities are indicated by black boxes and conservative changes are shaded. The positions of the eight signature motifs characteristic of SNF2 family proteins are indicated below the aligned sequences. Amino acid coordinates are indicated on the left; only the N-terminal 730 amino acids (of 1052 total) are shown for human SNF2h, though SEQ ID NO:5 shows the entire protein sequence. The deletion/frameshift caused by the *ddm1-2* allele occurs at amino acid 524. The *ddm1-6* frameshift occurs at amino acid 379, leading to translation of an additional 52 amino acids out of frame. The *ddm1-7* nonsense mutation

- 8 -

occurs at amino acid 549. Dashes indicate gaps introduced by the CLUSTAL W algorithm to maximize alignment (Thompson et al., Nucleic Acids Res. 22: 4673-4680, 1994). The alignment was processed by BOXSHADE v. 3.21.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Various terms relating to the biological molecules of the present invention are used throughout the specification and claims.

With reference to nucleic acids of the invention, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule.

With respect to RNA molecules of the invention the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

With respect to protein, the term "isolated protein" or "isolated and purified protein" is sometimes

- 9 -

used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein which has been sufficiently separated
5 from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid,
10 oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods,
15 agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

Nucleic acid sequences and amino acid sequences can be compared using computer programs that align the similar sequences of the nucleic or amino acids thus
20 define the differences. In the comparisons made in the present invention, the CLUSTLW program and parameters employed therein were utilized (Thompson et al., 1994, *supra*). However, equivalent alignments and similarity/identity assessments can be obtained through
25 the use of any standard alignment software. For instance, the GCG Wisconsin Package version 9.1, available from the Genetics Computer Group in Madison, Wisconsin, and the default parameters used (gap creation penalty=12, gap extension penalty=4) by that program may
30 also be used to compare sequence identity and similarity.

The term "substantially the same" refers to nucleic acid or amino acid sequences having sequence variation that do not materially affect the nature of the

- 10 -

protein (i.e. the structure, stability characteristics, substrate specificity and/or biological activity of the protein). With particular reference to nucleic acid sequences, the term "substantially the same" is intended to refer to the coding region and to conserved sequences governing expression, and refers primarily to degenerate codons encoding the same amino acid, or alternate codons encoding conservative substitute amino acids in the encoded polypeptide. With reference to amino acid sequences, the term "substantially the same" refers generally to conservative substitutions and/or variations in regions of the polypeptide not involved in determination of structure or function.

The terms "percent identical" and "percent similar" are also used herein in comparisons among amino acid and nucleic acid sequences. When referring to amino acid sequences, "percent identical" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical amino acids in the compared amino acid sequence by a sequence analysis program. "Percent similar" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical or conserved amino acids. Conserved amino acids are those which differ in structure but are similar in physical properties such that the exchange of one for another would not appreciably change the tertiary structure of the resulting protein. Conservative substitutions are defined in Taylor (1986, J. Theor. Biol. 119:205). When referring to nucleic acid molecules, "percent identical" refers to the percent of the nucleotides of the subject nucleic acid sequence that have been matched to identical nucleotides by a sequence analysis program.

- 11 -

With respect to antibodies, the term "immunologically specific" refers to antibodies that bind to one or more epitopes of a protein of interest, but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

With respect to oligonucleotides or other single-stranded nucleic acid molecules, the term "specifically hybridizing" refers to the association between two single-stranded nucleic acid molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

A "coding sequence" or "coding region" refers to a nucleic acid molecule having sequence information necessary to produce a gene product, when the sequence is expressed.

The term "operably linked" or "operably inserted" means that the regulatory sequences necessary for expression of the coding sequence are placed in a nucleic acid molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement other transcription control elements (e.g. enhancers) in an expression vector.

Transcriptional and translational control sequences are DNA regulatory sequences, such as

- 12 -

promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell. In particular, as used herein, the term "DNA transcriptional response element" refers to a DNA sequence specifically recognized for binding by a DNA binding protein characterized as a transcriptional regulator (either activator or suppressor).

The terms "promoter", "promoter region" or "promoter sequence" refer generally to transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. Typically, a promoter is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The typical 5' promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A "vector" is a replicon, such as plasmid, phage, cosmid, or virus to which another nucleic acid segment may be operably inserted so as to bring about the replication or expression of the segment.

The term "nucleic acid construct" or "DNA construct" is sometimes used to refer to a coding sequence or sequences operably linked to appropriate regulatory sequences and inserted into a vector for transforming a cell. This term may be used interchangeably with the term "transforming DNA". Such a

- 13 -

nucleic acid construct may contain a coding sequence for a gene product of interest, along with a selectable marker gene and/or a reporter gene.

5 The term "reporter gene" refers to genetic sequences which may be operably linked to a promoter region forming a transgene, such that expression of the reporter gene coding region is regulated by the promoter and expression of the transgene is readily assayed.

10 The term "selectable marker gene" refers to a gene product that when expressed confers a selectable phenotype, such as antibiotic resistance, on a transformed cell or plant.

The term "DNA construct" is sometimes used herein to refer to genetic sequence used to transform
15 plants and generate progeny transgenic plants. These constructs may be administered to plants in a viral or plasmid vector. Other methods of delivery such as Agrobacterium T-DNA mediated transformation and transformation using the biolistic process are also
20 contemplated to be within the scope of the present invention. The transforming DNA may be prepared according to standard protocols such as those set forth in "Current Protocols in Molecular Biology", eds. Frederick M. Ausubel et al., John Wiley & Sons, 1995.

25 A cell has been "transformed" or "transfected" by exogenous or heterologous DNA construct when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and plant
30 cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells

- 14 -

through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

II. Description of *DDM1* and its Encoded Polypeptide

In accordance with the present invention, a novel gene, *DDM1*, has been isolated from the flowering plant *Arabidopsis thaliana*. Through analysis of mutant plants, this gene has been identified as important for the maintenance of proper genomic cytosine methylation, and its function appears to be necessary to maintain gene silencing. Biochemical and molecular genetic results indicate that *DDM1* encodes a novel component of the DNA methylation machinery.

We have isolated the *DDM1* gene from *A. thaliana* using a map-based cloning approach, which is described in detail in Example 1 and shown in Figure 1. Briefly, the *DDM1* gene was initially localized to the bottom of the lower arm of chromosome 5 by reference to molecular markers segregating in an F2 family (parental cross: Columbia *ddm1/ddm1* X Landsberg erecta *DDM1/DDM1*). Next, recombination breakpoints in the region surrounding a *ddm1* mutation were isolated by collecting cross-over chromosomes by reference to flanking genetic markers. The recombination breakpoints delimited a region of approximately 30 kilobases. Cloned DNA corresponding to this genomic region was isolated by subcloning DNA from a

- 15 -

bacterial artificial chromosome (BAC) containing molecular markers mapping both proximal and distal to the *ddm1* marker. The nucleotide sequence of a single cosmid subclone encompassing the 30 kb region was determined to
5 identify six candidate genes, in addition to a tRNA gene and a previously identified succinate dehydrogenase structural gene.

The search for the *DDM1* gene focused on predicted genes 5 and 6, which fell in the center of the
10 genetic interval defined by recombination breakpoints with the *ddm1-2* marker. The *DDM1* gene was identified as predicted gene 6 based on DNA sequence alterations in four *ddm1* alleles (Figure 2). The EMS-generated *ddm1-2* mutation is a G to A transition in the splice donor site
15 of intron 11 that forces the use of an alternate splice donor site 56 bp upstream in exon 11 (Fig. 2B). The splicing defect leads to a deletion, a frameshift and premature translation termination upstream of predicted functional domains. The fast neutron-generated *ddm1-5*
20 (previously named *som8*; Mittelsten Scheid, O., Afsar, K. & Paszkowski, J. *Proc. Natl. Acad. Sci. USA* 95: 632-637, 1998).) allele contains an 82 bp insertion (1 bp deleted and replaced with 83 bp) in the second protein-coding exon, leading to an in-frame stop after 30 codons (15
25 wild-type codons plus 15 codons from the insertion). Premature translation termination is also predicted to result from two additional fast neutron alleles: *ddm1-6* (*som4*) corresponds to a frameshift (1 bp deletion) in exon 7 and *ddm1-7* (*som5*) is a nonsense mutation in exon
30 12. All four characterized *ddm1* alleles are expected to destroy or severely reduce gene function.

The wild-type *DDM1* gene encodes a predicted protein of 764 amino acids with a high degree of

- 16 -

similarity to SWI2/SNF2-like proteins. Members of the SWI2/SNF2 family are involved in various functions, including transcriptional co-activation, transcriptional co-repression, chromatin assembly and DNA repair.

5 Underlying these apparently diverse activities is the modification or disruption of protein-DNA interactions by multi-protein complexes which contain SWI2/SNF2-like components. Figure 3 shows an alignment among the deduced amino acid sequences of *A. thaliana* DDM1 and two
10 mammalian members of the SNF2 family, human SNF2h (SEQ ID NO:4; Arihara, T. et al., *Cytogenet. Cell Genet.* **81**, 191-193, 1998) and murine LSH (SEQ ID NO:5; lymphocyte specific helicase, LSH; Jarvis, C.D. et al. *Gene* **169**, 203-207, 1996). DDM1 contains the eight sequence motifs
15 diagnostic of SWI2/SNF2 family members (Bork, P. & Koonin, E.V. *Nucleic Acids Res.* **21**, 751-752, 1993). *A. thaliana* DDM1 and human SNF2h share 45 percent identity over the approximately 470 amino acid region comprising the signature motifs. Over a similar region, *A. thaliana*
20 DDM1 and murine LSH display approximately 50 percent identity, omitting the 47 residues (amino acids 276-322) apparently unique to LSH. Initial molecular phylogenetic analysis placed DDM1 in a small subfamily, within the SNF2 family, which contains proteins of unknown function,
25 including murine LSH (Eisen, J.A. et al. *Nucleic Acids Res.* **23**, 2715-2723, 1995). The proteins of known function most closely related to DDM1 are involved in chromatin remodeling and are grouped in the SNF2L/ISWI subfamily (Eisen et al., 1995, *supra*).

30 Without intending to be bound by any particular mechanism for the functionality of the *DDM1* gene product, analysis of the foregoing data indicates that the *DDM1* protein functions in the DNA methylation system by

- 17 -

affecting chromatin structure. Two general models for the *DDM1* action are envisioned. The *DDM1* protein may function as a transcriptional co-activator, similar to many SWI2/SNF2-like proteins, to increase the expression of a component of the DNA methylation system. *DDM1* does not affect DNA methyltransferase expression directly because *ddm1* mutant extracts contain wild-type methyltransferase activity (Kakutani et al., 1995, *supra*). However, an unidentified positive effector of DNA methylation may be a target. Alternatively, wild-type *DDM1* function may change chromatin structure to direct certain sequences to the methylation machinery or to facilitate the methylation of genomic substrates. The recently discovered interplay between cytosine methylation and histone acetylation, and the association of SWI2/SNF2-like proteins and histone deacetylases in chromatin remodeling complexes, makes it plausible that *DDM1* affects DNA methylation through modulation of histone modification or another aspect of chromatin structure. Another possibility is that *DDM1* plays a more direct role as a part of a nucleosome remodeling complex that increases the accessibility of the DNA methyltransferase to the hemimethylated substrates in newly replicated chromatin. The latter model is particularly attractive because it predicts that *ddm1* mutations will preferentially hypomethylate genomic sequences packaged in highly condensed chromatin while causing slower loss of methylation in more accessible sequences, consistent with the observed hypomethylation specificity of *ddm1* mutations. The isolation of the *Arabidopsis DDM1* gene in accordance with the present invention points to the importance of chromatin dynamics in the maintenance of cytosine methylation patterns and

- 18 -

identifies a novel component of the eukaryotic DNA methylation pathway.

A number of applications are contemplated for the novel gene of the invention and its encoded protein, and the discovery of the involvement of a *SWI2/SNF2*-like gene in the eucaryotic DNA methylation system. Such applications are described in greater detail below.

Although the *DDM1* genomic clone from *Arabidopsis thaliana* is described and exemplified herein, this invention is intended to encompass nucleic acid sequences and proteins from other organisms, including plants, yeast, insects and mammals, that are sufficiently similar to be used instead of the *Arabidopsis DDM1* nucleic acid and proteins for the purposes described below. These include, but are not limited to, allelic variants and natural mutants of SEQ ID NO:1, which are likely to be found in different species of plants or varieties of *Arabidopsis*. Because such variants are expected to possess certain differences in nucleotide and amino acid sequence, this invention provides an isolated *DDM1* nucleic acid molecule having at least about 60% (preferably 70% and more preferably over 80%) sequence homology in the coding regions with the nucleotide sequence set forth as SEQ ID NO:1 (and, most preferably, specifically comprising the coding region of SEQ ID NO:1). This invention also provides isolated polypeptide products of the open reading frames of SEQ ID NO:1, having at least about 60% (preferably 70% or 80% or greater) sequence homology with the amino acid sequences of SEQ ID NO:2. Because of the natural sequence variation likely to exist among *DDM1* genes, one skilled in the art would expect to find up to about 30-40% nucleotide sequence variation, while still maintaining

- 19 -

the unique properties of the *DDM1* gene and encoded polypeptide of the present invention. Such an expectation is due in part to the degeneracy of the genetic code, as well as to the known evolutionary success of conservative amino acid sequence variations, which do not appreciably alter the nature of the encoded protein. Accordingly, such variants are considered substantially the same as one another and are included within the scope of the present invention.

The following description sets forth the general procedures involved in practicing the present invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning procedures, such as those set forth in Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989) (hereinafter "Sambrook et al.") or Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1999) (hereinafter "Ausubel et al.") are used.

A. Preparation of *DDM1* Nucleic Acid Molecules, encoded Polypeptides and Antibodies Specific for the Polypeptides

1. Nucleic Acid Molecules

DDM1 nucleic acid molecules of the invention may be prepared by two general methods: (1) they may be synthesized from appropriate nucleotide triphosphates, or (2) they may be isolated from biological sources. Both methods utilize protocols well known in the art.

The availability of nucleotide sequence information, such as the cDNA having SEQ ID NO:1, enables preparation of an isolated nucleic acid molecule of the

- 20 -

invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramidite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. The resultant construct
5 may be purified according to methods known in the art, such as high performance liquid chromatography (HPLC). Long, double-stranded polynucleotides, such as a DNA molecule of the present invention, must be synthesized in stages, due to the size limitations inherent in current
10 oligonucleotide synthetic methods. Thus, for example, a long double-stranded molecule may be synthesized as several smaller segments of appropriate complementarity. Complementary segments thus produced may be annealed such that each segment possesses appropriate cohesive termini
15 for attachment of an adjacent segment. Adjacent segments may be ligated by annealing cohesive termini in the presence of DNA ligase to construct an entire long double-stranded molecule. A synthetic DNA molecule so constructed may then be cloned and amplified in an
20 appropriate vector.

DDM1 genes also may be isolated from appropriate biological sources using methods known in the art. In the exemplary embodiment of the invention, the *A. thaliana DDM1* clone was isolated from a BAC genomic
25 library of *A. thaliana*. In alternative embodiments, cDNA clones of *DDM1* may be isolated. A preferred means for isolating *DDM1* genes is PCR amplification using genomic templates and *DDM1*-specific primers.

In accordance with the present invention,
30 nucleic acids having the appropriate level sequence homology with part or all the coding regions of SEQ ID NO:1 may be identified by using hybridization and washing conditions of appropriate stringency. For example,

- 21 -

hybridizations may be performed, according to the method of Sambrook et al., using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 2X SSC and 0.1% SDS; (4) 2 hours at 45-55°C in 2X SSC and 0.1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology (Sambrook et al., 1989):

$$T_m = 81.5^{\circ}\text{C} + 16.6\text{Log} [\text{Na}^+] + 0.41(\% \text{G+C}) - 0.63 (\% \text{formamide}) - 600/\#\text{bp in duplex}$$

As an illustration of the above formula, using $[\text{Na}^+] = [0.368]$ and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C. Such a sequence would be considered substantially homologous to the sequences of the present invention.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a preferred embodiment, clones are maintained in plasmid cloning/expression vector, such as pGEM-T (Promega Biotech, Madison, WI) or pBluescript (Stratagene, La Jolla, CA), either of which is propagated in a suitable

- 22 -

E. coli host cell.

DDM1 nucleic acid molecules of the invention include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, this invention provides oligonucleotides (sense or antisense strands of DNA or RNA) having sequences capable of hybridizing with at least one sequence of a nucleic acid molecule of the present invention, such as selected segments of the DNA having SEQ ID NO:1. Such oligonucleotides are useful as probes for detecting *DDM1* genes or mRNA in test samples, e.g. by PCR amplification, or for the positive or negative regulation of expression of *DDM1* genes at or before translation of the mRNA into proteins.

The *DDM1* promoter and other expression regulatory sequences for *DDM1* are also expected to be useful in connection with the present invention. SEQ ID NO:1 shows about 550 bp of sequence upstream from the beginning of the coding region, which should contain such expression regulatory sequences. In addition, SEQ ID NO:3 constitutes about 5 kbp of additional upstream sequence, which should contain other regulatory sequences, such as enhancer elements.

25 2. Proteins

Polypeptides encoded by *DDM1* nucleic acids of the invention may be prepared in a variety of ways, according to known methods. If produced *in situ* the polypeptides may be purified from appropriate sources, e.g., plant parts.

Alternatively, the availability of nucleic acid molecules encoding the polypeptides enables production of the proteins using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into

- 23 -

an appropriate *in vitro* transcription vector, such a
pSP64 or pSP65 for *in vitro* transcription, followed by
cell-free translation in a suitable cell-free translation
system, such as wheat germ or rabbit reticulocytes. In
5 *in vitro* transcription and translation systems are
commercially available, e.g., from Promega Biotech,
Madison, Wisconsin or BRL, Rockville, Maryland.

According to a preferred embodiment, larger
quantities of DDM1-encoded polypeptide may be produced by
10 expression in a suitable procaryotic or eucaryotic
system. For example, part or all of a DNA molecule, such
as the coding portion of SEQ ID NO:1, may be inserted
into a plasmid vector adapted for expression in a
bacterial cell (such as *E. coli*) or a yeast cell (such as
15 *Saccharomyces cerevisiae*), or into a baculovirus vector
for expression in an insect cell. Such vectors comprise
the regulatory elements necessary for expression of the
DNA in the host cell, positioned in such a manner as to
permit expression of the DNA in the host cell. Such
20 regulatory elements required for expression include
promoter sequences, transcription initiation sequences
and, optionally, enhancer sequences.

The DDM1 polypeptide produced by gene
expression in a recombinant procaryotic or eucaryotic
25 system may be purified according to methods known in the
art. In a preferred embodiment, a commercially available
expression/secretion system can be used, whereby the
recombinant protein is expressed and thereafter secreted
from the host cell, to be easily purified from the
30 surrounding medium. If expression/secretion vectors are
not used, an alternative approach involves purifying the
recombinant protein by affinity separation, such as by
immunological interaction with antibodies that bind
specifically to the recombinant protein. Such methods
35 are commonly used by skilled practitioners.

- 24 -

The *DDM1*-encoded polypeptides of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures. Methods for analyzing the functional activity are available. For instance, DNA methylation levels are detectable by known methods. 5 Alternatively, the function of the *DDM1* gene product as part of a chromatin remodeling machine permits the use of *in vitro* assays for chromatin remodeling, which are known in the art (e.g., B.R. Cairns, *Trends in Biochem.* 23: 20- 10 25, 1998).

The present invention also provides antibodies capable of immunospecifically binding to polypeptides of the invention. Polyclonal or monoclonal antibodies directed toward the polypeptide encoded by *DDM1* may be 15 prepared according to standard methods. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. In a preferred embodiment, antibodies are prepared, which react immunospecifically with various epitopes of the 20 *DDM1*-encoded polypeptides.

**B. Uses of *DDM1* Nucleic Acids,
Encoded Proteins and Antibodies**

1. *DDM1* Nucleic Acids

25 *DDM1* nucleic acids may be used for a variety of purposes in accordance with the present invention. The DNA, RNA, or fragments thereof may be used as probes to detect the presence of and/or expression of *DDM1* genes. Methods in which *DDM1* nucleic acids may be utilized as 30 probes for such assays include, but are not limited to: (1) *in situ* hybridization; (2) Southern hybridization (3) northern hybridization; and (4) assorted amplification reactions such as polymerase chain reactions (PCR).

The *DDM1* nucleic acids of the invention may 35 also be utilized as probes to identify related genes from

- 25 -

other species, including but not limited to, plants, yeast, insects and mammals, including humans. As is well known in the art and described above, hybridization stringencies may be adjusted to allow hybridization of nucleic acid probes with complementary sequences of varying degrees of homology. Thus, *DDM1* nucleic acids may be used to advantage to identify and characterize other genes of varying degrees of relation to the exemplary coding sequence of SEQ ID NO:1, thereby enabling further characterization of this family of genes. Additionally, they may be used to identify genes encoding proteins that interact with protein encoded by *DDM1* (e.g., by the "interaction trap" technique).

As discussed above and in greater detail in Example 1, the similarity among plant *DDM1* and its *SWI2/SNF2* counterparts in yeast, *Drosophila* and mammals indicates that the functional aspects of these proteins will also be conserved. Thus, *DDM1* is expected to play an important role in DNA methylation and resultant down-regulation of gene expression. Plants engineered to over-express *DDM1* can be expected to have improved fidelity of the DNA methylation system. The evidence suggests that loss of *DDM1* function leads to reduction in the efficiency of maintenance methylation due to reduced accessibility of the methyltransferase enzyme to the substrate. Hence, excess *DDM1* function could lead to an increase in the fidelity of the inheritance of DNA methylation thereby reducing the occurrence of spurious methylation mistakes which could compromise the organism's viability or fecundity. In fact, there are experimental data demonstrating that loss of *DDM1* function leads to stochastic hypermethylation, and epigenetic lesion formation, as well. For these reasons, *DDM1* overexpression lines are expected to have useful properties.

- 26 -

Transgenic plants expressing the *DDM1* gene or antisense nucleotides can be generated using standard plant transformation methods known to those skilled in the art. These include, but are not limited to,

5 *Agrobacterium* vectors, PEG treatment of protoplasts, biolistic DNA delivery, UV laser microbeam, gemini virus vectors, calcium phosphate treatment of protoplasts, electroporation of isolated protoplasts, agitation of cell suspensions with microbeads coated with the

10 transforming DNA, direct DNA uptake, liposome-mediated DNA uptake, and the like. Such methods have been published in the art. See, e.g., Methods for Plant Molecular Biology (Weissbach & Weissbach, eds., 1988); Methods in Plant Molecular Biology (Schuler & Zielinski,

15 eds., 1989); Plant Molecular Biology Manual (Gelvin, Schilperoort, Verma, eds., 1993); and Methods in Plant Molecular Biology - A Laboratory Manual (Maliga, Klessig, Cashmore, Gruissem & Varner, eds., 1994).

The method of transformation depends upon the

20 plant to be transformed. The biolistic DNA delivery method is useful for nuclear transformation. In another embodiment of the invention, *Agrobacterium* vectors are used to advantage for efficient transformation of plant nuclei.

25 In a preferred embodiment, the gene is introduced into plant nuclei in *Agrobacterium* binary vectors. Such vectors include, but are not limited to, BIN19 (Bevan, 1984) and derivatives thereof, the pBI vector series (Jefferson et al., 1987), and binary

30 vectors pGA482 and pGA492 (An, 1986).

The *DDM1* gene may be placed under a powerful constitutive promoter, such as the Cauliflower Mosaic Virus (CaMV) 35S promoter or the figwort mosaic virus 35S promoter. Transgenic plants expressing the *DDM1* gene

35 under an inducible promoter (either its own promoter or a

- 27 -

heterologous promoter) are also contemplated to be within the scope of the present invention. Inducible plant promoters include the tetracycline repressor/operator controlled promoter.

5 Using an *Agrobacterium* binary vector system for transformation, the *DDM1* coding region, under control of a constitutive or inducible promoter as described above, is linked to a nuclear drug resistance marker, such as kanamycin resistance. *Agrobacterium*-mediated
10 transformation of plant nuclei is accomplished according to the following procedure:

(1) the gene is inserted into the selected *Agrobacterium* binary vector;

(2) transformation is accomplished by co-
15 cultivation of plant tissue (e.g., leaf discs) with a suspension of recombinant *Agrobacterium*, followed by incubation (e.g., two days) on growth medium in the absence of the drug used as the selective medium (see, e.g., Horsch et al. 1985);

20 (3) plant tissue is then transferred onto the selective medium to identify transformed tissue; and

(4) identified transformants are regenerated to intact plants.

It should be recognized that the amount of
25 expression, as well as the tissue specificity of expression of the *DDM1* gene in transformed plants can vary depending on the position of their insertion into the nuclear genome. Such position effects are well known in the art. For this reason, several nuclear
30 transformants should be regenerated and tested for expression of the transgene.

In some instances, it may be desirable to down-regulate or inhibit expression of endogenous *DDM1* in plants possessing the gene. One clear benefit to
35 engineering a reduction of *DDM1* function is to reduce

- 28 -

gene (including transgene) silencing. Plant lines with reduced or absent DDM1 function are expected to be viable based on results obtained with *Arabidopsis*. Further, it has been shown that gene silencing is suppressed in *ddm1* *Arabidopsis* lines (Jeddeloh et al., *Genes Devel.* 12:1714-1725, 1998). There are two other beneficial characteristics of DDM1 deficient plant lines. First, alteration in DNA methylation leads to changes in flowering time, and as such, is a potentially powerful tool for manipulating plant development. (See, e.g., Richards, *Trends in Genetics* 13: 319-323, 1998), Second, *ddm1* mutant lines exhibit inbreeding depression (a reduction in vigor after inbreeding) (Richards, *Trends in Genetics*, 1998, *supra*), a characteristic which may be desirable to include in situations where proprietary germplasms in hybrid plants are at risk of unauthorized use. For instance, a genetically engineered hybrid (containing one or more useful transgenes) could be further engineered to down-regulate endogenous DDM1 expression. Unauthorized inbreeding of such lines would be discouraged because the progeny of such lines would lack vigor.

To achieve the aforementioned benefits associated with reduced gene expression, DDM1 nucleic acid molecules, or fragments thereof, may also be utilized to control the production of DDM1-encoded proteins. In one embodiment, full-length DDM1 antisense molecules or antisense oligonucleotides, targeted to specific regions of DDM1-encoded RNA that are critical for translation, are used. The use of antisense molecules to decrease expression levels of a pre-determined gene is known in the art. In a preferred embodiment, antisense molecules are provided *in situ* by transforming plant cells with a DNA construct which, upon

- 29 -

transcription, produces the antisense sequences. Such constructs can be designed to produce full-length or partial antisense sequences.

5 In another embodiment, overexpression of *DDM1* is induced to generate a co-suppression effect. This excess expression serves to promote down-regulation of both endogenous and exogenous *DDM1* genes.

10 Optionally, transgenic plants can be created containing mutations in the region encoding the active site of *DDM1*. This embodiment may be preferred in certain instances.

From the foregoing discussion, it can be seen that *DDM1* and its homologs will be useful for introducing alterations in gene expression in an organism, for a
15 variety of purposes. As described above, for instance, the *Arabidopsis DDM1* gene can be used to isolate mutants or engineer organisms that express reduced function of *DDM1* orthologs. Based on results in *Arabidopsis*, such mutants or engineered organisms are expected to be viable
20 and display valuable characteristics, such as inbreeding depression and a reduction in gene silencing. In addition, we anticipate that dysfunction in human *DDM1* orthologs may contribute to diseases that involve alterations in DNA methylation, including cancer (Baylin,
25 S.B. et al., *Adv. Cancer Res.* 72: 141-196, 1998) and immunodeficiency/ chromosome instability/facial anomalies syndrome (ICF) (Smeets, D.F.C.M. et al., *Hum. Genet.* 94: 240-246, 1994).

30

2. *DDM1* Proteins and Antibodies

Purified *DDM1*-encoded proteins, or fragments thereof, may be used to produce polyclonal or monoclonal antibodies which also may serve as sensitive detection reagents for the presence and accumulation of *DDM1*-encoded

- 30 -

protein in cultured cells or tissues and in intact organisms. Recombinant techniques enable expression of fusion proteins containing part or all of the *DDM1*-encoded protein. The full length protein or fragments of the protein may be used to advantage to generate an array of monoclonal or polyclonal antibodies specific for various epitopes of the protein, thereby providing even greater sensitivity for detection of the protein in cells or tissue.

DDM1 gene products also may be useful as pharmaceutical agents if it is determined that *DDM1* loss of function plays a role in carcinogenesis, as mentioned above. The gene products could be administered as replacement therapy for persons having neoplasias associated with *DDM1* loss of function.

Polyclonal or monoclonal antibodies immunologically specific for *DDM1*-encoded proteins may be used in a variety of assays designed to detect and quantitate the protein. Such assays include, but are not limited to: (1) flow cytometric analysis; (2) immunochemical localization in cultured cells or tissues; and (3) immunoblot analysis (e.g., dot blot, Western blot) of extracts from various cells and tissues.

Polyclonal or monoclonal antibodies that immunospecifically interact with the polypeptide encoded by *DDM1* can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules.

The following specific examples are provided to illustrate embodiments of the invention. They are not intended to limit the scope of the invention in any way.

EXAMPLE 1
Map-Based Isolation of the
Arabidopsis thaliana DDM1 Gene

Construction of recombination breakpoint lines.

The recombination breakpoint lines were assembled in the
10 F3 generation from a parental cross between YI *DDM1*
ABA/YI *ddm1-2 ABA* (Columbia strain (Col)) and
yi DDM1 aba/*yi DDM1 aba* (Landsberg erecta strain
(La er)). The recessive *yi* mutation leads to a yellow
inflorescence. The recessive *aba* mutation causes a defect
15 in abscisic acid biosynthesis and a wilting phenotype.
Information on genetic markers and the *A. thaliana*
genetic map can be found at: [http://genome-](http://genome-www.stanford.edu/Arabidopsis/)
[www.stanford.edu/Arabidopsis/](http://genome-www.stanford.edu/Arabidopsis/). Selfed seeds from F1
YI ddm1-2 ABA/*yi DDM1 aba* plants were collected and 135-
20 F2 recombinants (*yi ABA*, yellow inflorescence, non-
wilting; or *YI aba*: green inflorescence, wilting) were
identified. Selfed seeds from 111 of the 135 recombinant
F2 individuals were planted to generate F3 tissue for
genomic DNA preparation. The genotype at the *DDM1* locus
25 was scored in the F3 generation by Southern blot analysis
using methylation-sensitive endonucleases as described
previously (Vongs, A., Kakutani, T., Martienssen, R.A. &
Richards, E.J. , *Science* **260**: 1926-1928, 1993).

Molecular markers. Two of the molecular
30 markers shown in Figure 1 were available from the
Arabidopsis research community: g4510 (*Arabidopsis*
Biological Resource Center (ABRC) stock# CD2-38) and
mi335 (ABRC stock# CD3-288). The remainder of the
molecular markers shown in Figure 1 were developed in
35 accordance with the present invention. ST10D21Bam is an

- 32 -

insert end subclone of the BAC (bacterial artificial chromosome) clone T10D21 constructed by complete cleavage with *Bam*HI and recircularization. sT10D21Bam recognizes a Col/La er *Pst*I RFLP (restriction fragment length polymorphism). Molecular marker A is an *Xba*I Col/La er RFLP marker recognized by a 5.7 kb *Hind*III fragment of the C38 cosmid insert. Marker B is a *Rsa*I Col/La er CAPS marker (Koneieczny & Ausubel, Plant J. 4: 403-410, 1993) (forward primer: 5'-TCAAGGAGATGATTCTGGGCGT-3', SEQ ID NO: 6; reverse primer: 5'-AAAGGACCCATTTACAGAACAC-3', SEQ ID NO:7). The remaining markers, C and D, correspond to RFLP's (*Bcl*I and *Pst*I, respectively) recognized by the succinate dehydrogenase cDNA clone, 105N23T7 (ABRC stock# 105N23T7).

Genomic library construction and screening. We screened the available *A. thaliana* BAC genomic libraries by standard colony hybridization techniques using radiolabeled 105N23T7 insert as a probe. The clone we subsequently focused upon, T10D21, came from the Texas A&M University BAC library (Choi et al., *Weeds World* 2: 17-20, 1995). To facilitate subsequent analysis, we cloned *Sau*3AI partially digested fragments from the T10D21 insert into the *Bam*HI site of SuperCos (Stratagene). We chose to further characterize one member of the resulting cosmid sublibrary, C38, which contained genetic markers that flanked *ddm1-2*. The C38 cosmid was submitted on April 20, 1999, under the provisions of the Budapest Treaty, with the American Type Culture Collection (Manassas VA), and assigned ATCC Accession No. 207208.

- 33 -

EXAMPLE 2

***DDM1 Gene Structure and Identification;
Sequence Determination of DDM1 Gene***

5 **DNA sequence determination.** C38 cosmid (~45 kb)
DNA, prepared using Qiagen columns and protocols, was
sonicated and 1-2 kb fragments isolated from a low-
melting temperature agarose gel. The size-selected DNA
was cloned into the *Sma*I site of a M13mp18 vector to
10 generate a shotgun library suitable for DNA sequence
determination. Single-stranded substrates were prepared
and sequenced using conventional dye-terminator cycle
sequencing protocols (Perkin-Elmer) on either an ABI 373
or ABI 377 automated DNA sequencer. The DNA sequence of
15 the *ddm1* alleles was determined using PCR-amplified
templates and oligonucleotide primers dispersed
throughout the *DDM1* gene. Sequence assembly and analysis
were accomplished using Phred/Phrap/Consed
(<http://www.mbt.washington.edu/>) and DNASTAR software
20 suites.

RT-PCR cDNA analysis. *DDM1* gene structure was
determined by analysis of the genomic DNA sequence and
the nucleotide sequence of RT-PCR (reverse transcription-
polymerase chain reaction) products encompassing the
25 coding region. *DDM1* and *ddm1-2* transcripts were analyzed
by RT-PCR as follows. Total RNA was prepared using the
Qiagen RNeasy™ protocol. Poly(A)+ transcripts were
collected on oligo-d(T)₂₅ magnetic Dynabeads (Dyna) and
first-strand cDNA synthesis performed following Dynal
30 protocols using Stratascript (Stratagene) reverse
transcriptase. Aliquots of the bead-immobilized first-
strand cDNA library were used as templates for PCR
amplification using KlenTaqI polymerase (Clontech). The
following oligonucleotide primers were used for the RT-
35 PCR experiment shown in Fig. 2b: forward,
5'-GCTGGAAGGGAAAGCTTAACAACC-3' (SEQ ID NO:8); reverse,

- 34 -

5'-ACACTGCCATCGATTCTGCAAACC-3' (SEQ ID NO:9).

GenBank accession numbers and SEQ ID NOS.

Arabidopsis DDM1 genomic DNA sequence: SEQ ID NO:1;

Arabidopsis DDM1 deduced amino acid sequence: SEQ ID NO:2;

5 *Arabidopsis DDM1* 5' upstream genomic DNA sequence: SEQ ID NO:3;

Mus musculus lymphocyte specific helicase (LSH); Genbank Accession No. AAB08015; SEQ ID NO:4;

10 *Homo sapiens* SNF2h; Genbank Accession No. AB010882; SEQ ID NO:5;

succinate dehydrogenase cDNA 105N23T7, T22529;

primers: SEQ ID NOS: 6-9.

15 While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set
20 forth in the following claims.

- 35 -

SEQUENCE LISTING

<110> Eric J. Richards
Jeffrey A. Jeddelloh

<120> Plant Gene that Regulates DNA
Methylation

<130> WashU CI-0014PCT

<150> US 60/_____
<151> 1998-04-30

<150> US 09/104,070
<151> 1998-06-24

<160> 9

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 5000
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (535)...(566)

<221> CDS
<222> (772)...(850)

<221> CDS
<222> (986)...(1252)

<221> CDS
<222> (1354)...(1440)

<221> CDS
<222> (1549)...(1895)

<221> CDS
<222> (1976)...(2165)

<221> CDS
<222> (2251)...(2426)

<221> CDS
<222> (2559)...(2625)

<221> CDS
<222> (2703)...(2892)

<221> CDS
<222> (2975)...(3070)

<221> CDS
<222> (3148)...(3242)

<221> CDS
<222> (3317)...(3436)

- 36 -

```

<221> CDS
<222> (3540)...(3659)

<221> CDS
<222> (3745)...(3843)

<221> CDS
<222> (3934)...(4038)

<221> CDS
<222> (4130)...(4354)

<221> gene
<222> (535)...(4354)
<223> /gene="DDM1"

<221> mutation
<222> (785)...(785)
<223> /note= "site of ddm1-5 (som8) mutation; delete G at 785 and
replace with 82 bp"

<221> mutation
<222> (2384)...(2385)
<223> /note= "site of ddm1-6 (som4) mutation; delete G at 2384 or
2385"

<221> misc_feature
<222> (3186)...(3186)
<223> /note= "alternate splice donor site used in ddm1-2"

<221> mutation
<222> (3243)...(3243)
<223> /note= "site of ddm1-2 mutation; G to A"

<221> mutation
<222> (3337)...(3337)
<223> /note= "site of ddm1-7 (som5) mutation; G to A"

<221> tRNA
<222> (4755)...(4826)
<223> /note= "complement of predicted tRNA-glu"

<400> 1
tgatcatttt cttcctccgg ccaatttgca gatcgaaaaa tgatttagct ttttattaaa      60
aatattgtta ttcgttttta gccgatatca taactttttg agatacatta tcaacacact      120
cgtgcaactg agatattctt gacacaattt ttgcatttga aattggcaat tttgtactac      180
tcatatagtt tgaagcttca attcactaca aaggttatta ctaatttgtg cgacaaatcc      240
agcagattta ataatgccca ttccattaaa tgtttttttag ttttaataata ggatgatcat      300
atgacccaaa tcgtaaataa gggttagggg taaacctgtc atttcaagct tcccgcccat      360
gggcgctact cccaatttaa taaaaaataa gaaaataggc gtaaatatga gagtgtgttt      420
tttcaatata ccctcggttt tgaatttgct ctcaaaagcg acggagacga ctgtttggct      480
cgggtgatttc tcccgcggtt tgggtttttc ttaccggaat ttctttctcc ttcgatggtt      540
agtctgcgct ccagaaaagt tattccgtaa gtccttcac ctttctttt catttcgta      600
tttcgcggca ttttctaggt ccttaacgct ctcgaaatcg ctcgctgttc ttggtggttt      660
ttggttccct ctctgcgtaa ttttgtttgt cgtgtttttg gattatattc tctgactatt      720
gggtctcactg ttgatttatc atttctcgat ttgggatttt tggactctta gggcttcgga      780
aatgggtcagc gacgggaaaa cggagaaaga tgcgtctggt gattcaccoca cttctgttct      840
caacgaagag gtttgttcta tgttctacta ttttgccctt gtagtgtggt tgctttgtga      900
aactttgtgt gttactcttt gtttctttta atctgggggtg ttctgtaaat gggtcctttt      960
tggtcctttt tttctgaatg tgaaggaaaa ctgtgaggag aaaagtgtta ctgttgtaga      1020
ggaagagata cttctagcca aaaatggaga ttcttctctt atttctgaag ccatggctca      1080
ggaggaagag cagctgctca aacttcggga agatgaagag aaagctaaca atgctggatc      1140

```

tgctgttgc	cctaactctga	atgaaactca	gtttactaaa	cttgatgagc	tcttgacgca	1200
aactcagctc	tactctgagt	ttctccttga	gaaaatggag	gatatcacia	ttgtaactctt	1260
ctttatttct	ttcttctttg	tggtttctca	cttttcgaat	gggagtcatt	attcttagtt	1320
tgaacaactt	gtgggtgaaa	tttgttttgc	tagaatggga	tagaaagtga	gagccaaaaa	1380
gctgagcccg	agaagactgg	tcgtggacgc	aaaagaaagg	ctgcttctca	gtacaacaat	1440
gttgggtcca	tttatataat	tttcaactac	tatgcatgat	cttgatatata	ttgttttttc	1500
tgcttgtttg	agaaagtaac	ttacttggat	gcttttttct	tcaatcagac	taagggtctaa	1560
agagcgggtg	ctgctatgat	ttcaagatct	aaagaagatg	gtgagaccat	caactcagat	1620
ctgacagagg	aagaaacagt	catcaaactg	cagaatgaac	tttgtcctct	tctcactggg	1680
ggacagttaa	agtcttatca	gcttaaagg	gtcaaattggc	taatatcatt	gtggcagaat	1740
ggtttgaatg	gaatatttagc	tgatcaaattg	ggacttggaa	agacgattca	aacgatcggt	1800
ttcttatcac	atctgaaagg	gaatgggttg	gatgggtccat	atctagtcac	tgctccactg	1860
tctacacttt	caaattgggt	caatgagatt	gctaggtact	ctcatggcca	tatgtgtttg	1920
tatagatcca	atgcttttggg	gtttctgttg	aaagttttct	taccttttcc	attaggttca	1980
cgccttccat	caatgcaatc	atctaccatg	gggataaaaa	tcaaagggat	gagctcagga	2040
ggaagcacat	gcctaaaact	gttgggtccca	agttccctat	agttattact	tcttatgagg	2100
ttgccatgaa	tgatgctaaa	agaattctgc	ggcactatcc	atggaaatat	gttgtgattg	2160
atgaggtaaa	ttccgagatt	ggtcaatgta	ctaggctttg	aagatcaaga	tgatctctct	2220
aactgataat	tttgttcttg	tatatattag	ggccacaggt	tgaaaaacca	caagtgtaaa	2280
ttgttgaggg	aactaaaaca	cttgaagatg	gataacaaac	ttctgctgac	aggaacacct	2340
ctgcaaaaata	atctttctga	gctttgggtct	ttgttaaatt	ttattctgcc	tgacatcttt	2400
acatcacatg	atgaatttga	atcatgggtac	aaacatgggtc	cttttctact	attatcccta	2460
actagtcttc	tttttttttc	tttttttgtt	aacactgggtg	gcagcttttt	gacattttatt	2520
cctttcttag	tatctaactg	atagatgagt	ctctacaggt	ttgatttttc	tgaaaagaac	2580
aaaaacgaag	caaccaagga	agaagaagag	aaaagaagag	ctcaagtatg	tacaattata	2640
tcaattttcc	tttattttct	tgattgtatt	tatgtcttat	gctaagggtg	catcttgtct	2700
agggtgtttc	caaacttcat	ggtatactac	gaccattcat	ccttcgaaga	atgaaatgtg	2760
atgttgagct	ctcacttcca	cggaaaaagg	agattataat	gtatgctaca	atgactgac	2820
atcagaaaaa	gttccaggaa	catctgggtg	ataacacgtt	ggaagcacat	cttggagaga	2880
atgccatccg	aggtacatga	tctatttttt	ttttttaata	ctttgtttaa	ttatgtcatt	2940
ttctgcattg	atltgttcat	ccctataact	tcagggtcaag	gctggaaagg	aaagcttaac	3000
aacctgggtc	ttcaacttcg	aaagaactgc	aaccatcctg	accttctcca	ggggcaataa	3060
gatgggtcat	gtatgtcagt	ttcttttaag	aaacgtaaga	aaaacttctg	tcatactgtt	3120
ctgtctaatt	gtttcatttc	gtgacagatc	tctaccctcc	tggtgaagag	attgttggac	3180
agtgtggtaa	attccgctta	ttggagagat	tacttgttcc	ggtattttgcc	aataatcaca	3240
aagtatgttt	cacaaaccca	tggtcgttag	ctcattttccc	tttgagaact	tctctgatcc	3300
atttgctgat	gaccagggtcc	ttatctttctc	ccaatggacg	aaacttttgg	acattatgga	3360
ttactacttc	agtgagaagg	ggtttgaggt	ttgcagaatc	gatggcagtg	tgaagctgga	3420
tgaaaggaga	agacagggttt	cacctgtgct	tatgctgctt	ttgcgttgct	tttaagcaat	3480
attctgacca	aattattata	ccataaggct	tctctctctc	tctcttttgc	ttgaaacaga	3540
ttaaagattt	cagtgatgag	aagagcagct	gtagatatatt	tctcctgagt	accagagctg	3600
gaggactcgg	aatcaatctt	actgctgctg	atacatgac	cctctatgac	agcgactggg	3660
taatcaaatc	aattaattta	ttttctttga	aggaaaatct	ttctctttcg	tggtgtctcc	3720
aactgtgttt	tgtctgatct	ccagaaccct	caaattggact	tgcaagccat	ggacagatgc	3780
cacagaatcg	ggcagacgaa	acctgttcat	gttttataggc	tttccacggc	tcagtcgata	3840
gaggtaaaaac	tctttgttgt	tcatatcaat	caatcttaac	ttcaaaccat	tgagattggt	3900
gcctcatgag	attgggtttat	gacatttgc	cagaccggg	ttctgaaacg	agcgtacagt	3960
aagctcaagc	tggaacatgt	ggttattggc	caagggcag	ttcatcaaga	acgtgccaa	4020
tcttcaacac	cttttagaggt	tttaacttct	cttaaagctc	aatccttttt	agatacactt	4080
attatcaaca	aaatctccta	ttgacagctt	gaaccaaaact	aacacacagg	aagaggacat	4140
actggcggtg	cttaagggaag	atgaaactgc	tgaagataag	ttgatacaaa	ccgatataag	4200
cgatgcggat	cttgacaggt	tacttgaccg	gagtgacctg	acaattactg	caccgggaga	4260
gacacaagct	gctgaagctt	ttccagtgaa	gggtccaggt	tgggaaagtgg	tccctgctag	4320
ttcggggagga	atgctgtctt	ccctgaacag	ttaggacaca	ttaataagcc	aggccttgaa	4380
accacttctg	tggttttttt	ttttttttcc	ggaacatgat	cggttacttt	tggtctggag	4440
gattttaatta	ttagaggggt	cggaagtttt	tgtaagttaa	agaactcact	taaaaccctg	4500
aaaacatgac	agttaatgg	gatttagctct	caatgtgatg	aaaacaattg	gccctctgat	4560
tttgctgttg	cggtaatat	atgacttgtg	tacgtttata	gtctttgtag	tctgcaattt	4620
tggtcattgag	ctattttctca	cgaacttatg	ggatcttatg	ttttggattt	gggatttgg	4680
aacttatatg	attagggtca	atagtttcac	agaatattaa	aaacttgagt	aggggttaaa	4740
aaagaagcaa	aaagctccga	tgccggggaat	cgaaccggg	tctcctgggt	gaaagccaga	4800
tatcctaacc	gctggacgac	atcggatttg	ttgatgtcta	ttcttgtaaa	tagtaaatat	4860

- 38 -

```

ttagtttttat cggtttttgca tctaattggac taaaacatga acacgagacg cgcacaagaa 4920
tgaatggggc aggcacacaaa catttgggta aaagtatgca gtgggggtatt attgacaatt 4980
tgaccattac aagagctaata                                     5000

```

<210> 2
 <211> 764
 <212> PRT
 <213> Arabidopsis thaliana

<400> 2

Met	Val	Ser	Leu	Arg	Ser	Arg	Lys	Val	Ile	Pro	Ala	Ser	Glu	Met	Val
1				5					10					15	
Ser	Asp	Gly	Lys	Thr	Glu	Lys	Asp	Ala	Ser	Gly	Asp	Ser	Pro	Thr	Ser
			20					25					30		
Val	Leu	Asn	Glu	Glu	Glu	Asn	Cys	Glu	Glu	Lys	Ser	Val	Thr	Val	Val
		35					40					45			
Glu	Glu	Glu	Ile	Leu	Leu	Ala	Lys	Asn	Gly	Asp	Ser	Ser	Leu	Ile	Ser
	50					55					60				
Glu	Ala	Met	Ala	Gln	Glu	Glu	Gln	Leu	Leu	Lys	Leu	Arg	Glu	Asp	
65				70					75					80	
Glu	Glu	Lys	Ala	Asn	Asn	Ala	Gly	Ser	Ala	Val	Ala	Pro	Asn	Leu	Asn
				85					90				95		
Glu	Thr	Gln	Phe	Thr	Lys	Leu	Asp	Glu	Leu	Leu	Thr	Gln	Thr	Gln	Leu
			100					105					110		
Tyr	Ser	Glu	Phe	Leu	Leu	Glu	Lys	Met	Glu	Asp	Ile	Thr	Ile	Asn	Gly
		115					120					125			
Ile	Glu	Ser	Glu	Ser	Gln	Lys	Ala	Glu	Pro	Glu	Lys	Thr	Gly	Arg	Gly
	130					135					140				
Arg	Lys	Arg	Lys	Ala	Ala	Ser	Gln	Tyr	Asn	Asn	Thr	Lys	Ala	Lys	Arg
145					150					155					160
Ala	Val	Ala	Ala	Met	Ile	Ser	Arg	Ser	Lys	Glu	Asp	Gly	Glu	Thr	Ile
				165					170					175	
Asn	Ser	Asp	Leu	Thr	Glu	Glu	Glu	Thr	Val	Ile	Lys	Leu	Gln	Asn	Glu
			180					185					190		
Leu	Cys	Pro	Leu	Leu	Thr	Gly	Gly	Gln	Leu	Lys	Ser	Tyr	Gln	Leu	Lys
		195					200					205			
Gly	Val	Lys	Trp	Leu	Ile	Ser	Leu	Trp	Gln	Asn	Gly	Leu	Asn	Gly	Ile
	210					215					220				
Leu	Ala	Asp	Gln	Met	Gly	Leu	Gly	Lys	Thr	Ile	Gln	Thr	Ile	Gly	Phe
225					230					235					240
Leu	Ser	His	Leu	Lys	Gly	Asn	Gly	Leu	Asp	Gly	Pro	Tyr	Leu	Val	Ile
			245						250					255	
Ala	Pro	Leu	Ser	Thr	Leu	Ser	Asn	Trp	Phe	Asn	Glu	Ile	Ala	Arg	Phe
			260				265						270		
Thr	Pro	Ser	Ile	Asn	Ala	Ile	Ile	Tyr	His	Gly	Asp	Lys	Asn	Gln	Arg
		275					280					285			
Asp	Glu	Leu	Arg	Arg	Lys	His	Met	Pro	Lys	Thr	Val	Gly	Pro	Lys	Phe
	290					295					300				
Pro	Ile	Val	Ile	Thr	Ser	Tyr	Glu	Val	Ala	Met	Asn	Asp	Ala	Lys	Arg
305					310					315					320
Ile	Leu	Arg	His	Tyr	Pro	Trp	Lys	Tyr	Val	Val	Ile	Asp	Glu	Gly	His
			325						330					335	
Arg	Leu	Lys	Asn	His	Lys	Cys	Lys	Leu	Leu	Arg	Glu	Leu	Lys	His	Leu
			340					345					350		
Lys	Met	Asp	Asn	Lys	Leu	Leu	Leu	Thr	Gly	Thr	Pro	Leu	Gln	Asn	Asn
		355					360					365			
Leu	Ser	Glu	Leu	Trp	Ser	Leu	Leu	Asn	Phe	Ile	Leu	Pro	Asp	Ile	Phe
	370					375					380				
Thr	Ser	His	Asp	Glu	Phe	Glu	Ser	Trp	Phe	Asp	Phe	Ser	Glu	Lys	Asn
385					390					395					400
Lys	Asn	Glu	Ala	Thr	Lys	Glu	Glu	Glu	Glu	Lys	Arg	Arg	Ala	Gln	Val
				405					410					415	

- 39 -

Val Ser Lys Leu His Gly Ile Leu Arg Pro Phe Ile Leu Arg Arg Met
 420 425 430
 Lys Cys Asp Val Glu Leu Ser Leu Pro Arg Lys Lys Glu Ile Ile Met
 435 440 445
 Tyr Ala Thr Met Thr Asp His Gln Lys Lys Phe Gln Glu His Leu Val
 450 455 460
 Asn Asn Thr Leu Glu Ala His Leu Gly Glu Asn Ala Ile Arg Gly Gln
 465 470 475 480
 Gly Trp Lys Gly Lys Leu Asn Asn Leu Val Ile Gln Leu Arg Lys Asn
 485 490 495
 Cys Asn His Pro Asp Leu Leu Gln Gly Gln Ile Asp Gly Ser Tyr Leu
 500 505 510
 Tyr Pro Pro Val Glu Glu Ile Val Gly Gln Cys Gly Lys Phe Arg Leu
 515 520 525
 Leu Glu Arg Leu Leu Val Arg Leu Phe Ala Asn Asn His Lys Val Leu
 530 535 540
 Ile Phe Ser Gln Trp Thr Lys Leu Leu Asp Ile Met Asp Tyr Tyr Phe
 545 550 555 560
 Ser Glu Lys Gly Phe Glu Val Cys Arg Ile Asp Gly Ser Val Lys Leu
 565 570 575
 Asp Glu Arg Arg Arg Gln Ile Lys Asp Phe Ser Asp Glu Lys Ser Ser
 580 585 590
 Cys Ser Ile Phe Leu Leu Ser Thr Arg Ala Gly Gly Leu Gly Ile Asn
 595 600 605
 Leu Thr Ala Ala Asp Thr Cys Ile Leu Tyr Asp Ser Asp Trp Asn Pro
 610 615 620
 Gln Met Asp Leu Gln Ala Met Asp Arg Cys His Arg Ile Gly Gln Thr
 625 630 635 640
 Lys Pro Val His Val Tyr Arg Leu Ser Thr Ala Gln Ser Ile Glu Thr
 645 650 655
 Arg Val Leu Lys Arg Ala Tyr Ser Lys Leu Lys Leu Glu His Val Val
 660 665 670
 Ile Gly Gln Gly Gln Phe His Gln Glu Arg Ala Lys Ser Ser Thr Pro
 675 680 685
 Leu Glu Glu Glu Asp Ile Leu Ala Leu Leu Lys Glu Asp Glu Thr Ala
 690 695 700
 Glu Asp Lys Leu Ile Gln Thr Asp Ile Ser Asp Ala Asp Leu Asp Arg
 705 710 715 720
 Leu Leu Asp Arg Ser Asp Leu Thr Ile Thr Ala Pro Gly Glu Thr Gln
 725 730 735
 Ala Ala Glu Ala Phe Pro Val Lys Gly Pro Gly Trp Glu Val Val Leu
 740 745 750
 Pro Ser Ser Gly Gly Met Leu Ser Ser Leu Asn Ser
 755 760

<210> 3

<211> 5000

<212> DNA

<213> Arabidopsis thaliana

<400> 3

tgctgaagtt	tccatggaag	attgtgacca	cgacgatgaa	gctgaagatt	ctgggtcacgt	60
tgaaaacctt	tgttacagat	ttcgcaaacg	aatcgattcg	ttgccataag	tgtttttaggt	120
gacaaagcta	tcacttcagc	gtctggatct	gaatttagac	aatcagtgag	aacaactaaa	180
aacagaaaat	ttcaaaactca	aaaaacagaa	aaaaaaaaagt	ttggattttt	gagaagtacc	240
aggcattcca	ggaagattcc	gtttcttctt	cccgcaggat	ttaggagtta	gatttttggtt	300
tccggctgat	gagacgcttg	catcgccgga	aactgtagag	gaattatcta	aatcaaccgg	360
catgtttcaa	agatactaaa	ttccaatctt	tgaacacaaa	aaggaagaag	caaattctcag	420
ctcagctcaa	tctagggttt	atcatcctcc	tctactctct	tttagtctct	ctttctctct	480
ctcttcttca	gctaccagtc	aatctgcttt	tcgtaaaaat	ctccttttcc	cctttccgcc	540
accaaacttt	tctgataact	cactctctga	cctctcttct	tcaaaaagat	ttaaaacccc	600
caaaagaaaa	agaaaaaaa	tcaaaacttc	attacccaag	aaatctctta	atcatttaac	660

ccagactctt	tcttctccac	acgcactctt	tatccaccgt	ccaccgatct	gatccaacgg	720
ctgagatttc	accggagacg	agttatctct	actacttccg	gcttgtttct	ctctgaagaa	780
tcaccggaaa	aaaaaataag	gcggcttggt	tgtgagactt	tgtgtgaaag	cttcaacctt	840
ttttttcttt	ttctttgggt	tgtccaagaa	aaaggagcct	tcttcttctt	ttctctctct	900
ggagacaatt	atactaattt	ttttcttttc	aactttttcac	cctttttttt	ttgttaacaa	960
acatttttta	tacataattg	tgtcgacttt	caagttccaa	gtatctaaat	ctgtattttg	1020
gactcccatg	caaataatta	aaatagaata	atctttttgt	agatttttaa	ttgaaaacgg	1080
tgtagaaagg	ttaaaagcac	caaacaaaac	gagtaaatag	atattgtaat	aattttttca	1140
cctttatgga	aaagattata	tcatagacga	tgtacacaga	tgaaaattag	aaaatggcat	1200
gtgaatatat	gcagtaccca	atgaatgcaa	tatcaggttt	gtattatttt	tctattgtat	1260
ctctacatgt	tacgtaatca	aacgatcaag	taatttatta	atattgtcga	tggcgtagaa	1320
attataaatt	tattttatgt	cattgtttac	tatatagatt	ttgagctaaa	cgacttattt	1380
tgtcaaaaga	tatatccgtg	tttggtttta	gattgggttt	tagtattttc	aatattaatc	1440
taaattctta	gcttatgaac	atgtcaataa	acaaaaaaat	tattttactg	tcactgtcct	1500
tagacgggga	caaaggaggg	tattaccgtc	gcgttgctcg	accgtaaaat	aaattaaccaa	1560
attttgttgt	tgaacgaata	acatttttta	ctgtgggaat	ttgtcgtgta	gcattacggt	1620
cgaaatcgca	atttgttttc	ttctttgtgg	gtgtatatatt	ctgggttaacg	aaactataac	1680
ccaattttaat	gcaatgttcg	tctgtttttg	ttgactttga	cccttttttg	gtaatatctg	1740
ttcagctttt	gttttaacgt	tttcattgcc	ttgtaggcat	ctgagaagct	cagattctga	1800
cacgtgtctt	ttgttatctg	aatttgcata	cgttggataa	acatgacgct	gacaggtgga	1860
ttgaaaagta	accagcttgg	atttctgtgt	atatgttaca	ccgccacttc	ccttaatttc	1920
ttcgttctta	gttaaaataa	aaaagggtta	atttatgagt	aaaagtatgt	aaaacgacaa	1980
cgattactat	agaattataa	atttatcttt	gcttagtaat	ttgcacttaa	gattggattc	2040
aaattttgta	aaaagcgaat	gttacatata	tgtccattga	aaaaattgca	tttgacttta	2100
caagcatgtg	aattaattaa	tttgggaccc	ctttttttgt	tagtttcaaa	ggaagaatta	2160
ttttaggctg	agatgggtcc	ctccataaac	tcactattct	gccagcatac	aaattcctta	2220
acatatgggtc	caaatagcag	ttccaaccac	tagtatccaa	taataatctg	aacaaattat	2280
ctttcttttt	ttttcctgat	aatcttgtat	ttgtttgttc	aatgagctta	atacgtatat	2340
tagttatgac	ttataactaa	atactttgac	tcacttgatc	cgtacacatt	gatttcgttt	2400
attcaaatcc	gaacaacgta	atgatctttt	tggggccgagt	tatttgtatt	ctcaacctga	2460
gtccaacctat	gctttatggg	cttttctgtt	tatttatgca	tgtaaagtta	tttaagtctg	2520
caaataacca	catattgtat	gaatgtaatt	actatgattt	aagggcactg	cttttctgtt	2580
ttcacgttgt	tttcgaaatt	gctattgcgt	gtgatatctg	tgttggacca	attattgaaa	2640
aggacaaggc	tgactctggg	ttttaatgag	tagtcccat	gggagttatg	ttcattttacc	2700
acacattttt	ttgtatagta	tagtatgagt	ttttatattga	tatcttttat	cttcggaaaa	2760
taaatgggtc	aaattgtttg	tctaaaaatg	cacacatgaa	tatcttgggt	tctcacacaa	2820
ttgtaggaaa	caaattaata	tttgttgcga	aaataatggt	attattttat	catacgaatt	2880
cctagagaaa	atgggtggcaa	aagaggcaaa	gactaaacta	atgaatttaa	aatatgaaaa	2940
tgatggaatg	actgggtttac	caatattaca	gtatattgta	attttataaa	aacgaatcct	3000
gaagaagagg	gcaaacccca	agaccacgca	aatcagttca	caaataatgaa	aattttccaat	3060
aactagaaaa	acatgtgcac	ttatcttttt	ccatcattcg	gattttttaca	atggaaaattt	3120
tgaccatgga	gcgcaagtgt	tatagtattt	tattattaat	caatattaat	atcattatttc	3180
ggatccatgc	attctatata	actatgtcca	ccatcttact	tgtgtctatg	ttgcaacttc	3240
aacgtcgtat	atataatagg	attgttgtca	cgaatacaat	gctaatttaag	gaagatttgtg	3300
acttctcgga	aaatttagaa	ctaatttaaga	gtggaaactaa	aatgccaatg	aaaatagcct	3360
aaatcaaagg	agaaccacaa	atataaatgt	gaagacctta	aaaaacaatt	aaacgaggac	3420
gaaacaaatt	ttggaatcat	caattatacg	aaaaaaagaa	gaaagaaaaa	agagggtttca	3480
tgaatcacag	tagtgctgac	aatcttcgaa	ccatttgggtg	gtttcatata	atcgatcacc	3540
aatagaacaa	aagagaaaca	gaggaacaga	aagaatagaa	ggagtgggaa	gtgtatgagg	3600
aagctgtgtc	cgaacataga	caaagacgat	ggtctggaga	cggtgttgga	agttccgata	3660
ccggaggaga	tgttttccgg	tatgggcaac	aacgttgcac	ttaggtggca	aaatatgatg	3720
acgtggatga	aagctcaaac	gtctgataaa	tggtcgcaac	cgcttatcgc	cgctcgtatc	3780
aacgagctcc	ggttccttct	ctacctcgtt	ggctcgcctc	ttatacctct	ccagggttcaa	3840
gtcggctcact	ctgttcataa	gcccgtcaaa	gattgtccca	ttgtaagtca	ttcaaaatca	3900
atccttatga	aaacataaca	aagatgttga	aaatatgatt	cctctttttt	ttttcttttt	3960
ttcttttatg	atcaaaaccc	aaaaaaagtca	ttaccctgct	tcgtaagtat	tcaacataaa	4020
gttggttaatc	catgtgttgt	actctgcaag	tctgcattac	attattcatc	gtacacagag	4080
tcaccaaact	cagtttcatt	gtttttttgc	ttatgaatta	cgattgcagc	aagcttcaac	4140
ggcgaaatag	attgtacagc	agtacatagc	agcgacggga	ggaccacagg	gcttaaacgc	4200
cgtgaacagc	atgtgcgtca	cgggacaagt	gaagatgacg	gcgtcggagt	ttcatcaagg	4260
agatgattcg	ggcgttaatc	taaagagcaa	cgacgaaatg	gggtggtttcg	ttttatggca	4320
aaaggatcca	gatcttttgt	gtttggagct	cgctcgtctc	gggtgcaaa	tggatatgtg	4380

- 41 -

```

gaagcaacgg tcgggtttca tggcgacatt cctctaacca gcaaactccg gcgtctacsg 4440
gaacgccaar acctctccgc cggtttwtac aggtccaatc cggttattga ttttttttk 4500
gatgtaatgt ccggtttctca aaatgttgaa ccggtgggtt atttattgtt tggagcaggg 4560
gttaratcct cgttcgacgg cgaatctgtt tcttgacgca aacgtgtatc ggagagaaga 4620
taatcaacgg tgaggattgc tttatcttga aactggagac gagtccggcg gttcagaag 4680
ctcaaagcgg tccgaatttt gagataattc atcacacgat atgggggttat tttagtcaaa 4740
gatcgggact tttgattcag ttcgaagatt cgcggctttt gagaatgagg accaaggaag 4800
acgaagatgt cttctgggag actagtgtg agtcgggtgat ggatgattac cgatacgttg 4860
acaatgtgaa catcgctcac ggcgggaaaa catcggtcac ggttttccgg tacggtgaag 4920
cgtcggcgaa tcatcggaga cagatgacgg agaagtggag gatagaagaa gttgatttta 4980
atgtttgggg tctctccgtt 5000

```

<210> 4
 <211> 603
 <212> PRT
 <213> Mus musculus

<400> 4

Met	Leu	Trp	Glu	Asn	Gly	Ile	Asn	Gly	Ile	Leu	Ala	Asp	Glu	Met	Gly
1				5					10				15		
Leu	Gly	Lys	Thr	Val	Gln	Cys	Ile	Ala	Thr	Ile	Ala	Leu	Met	Ile	Gln
			20					25				30			
Arg	Gly	Val	Pro	Gly	Pro	Phe	Leu	Val	Cys	Gly	Pro	Leu	Ser	Thr	Leu
		35					40				45				
Pro	Asn	Trp	Met	Ala	Glu	Phe	Lys	Arg	Phe	Thr	Pro	Glu	Ile	Pro	Thr
	50					55					60				
Leu	Leu	Tyr	His	Gly	Thr	Arg	Glu	Asp	Arg	Arg	Lys	Leu	Val	Lys	Asn
65					70					75					80
Ile	His	Lys	Arg	Gln	Gly	Thr	Leu	Gln	Ile	His	Pro	Val	Val	Val	Thr
				85					90					95	
Ser	Phe	Glu	Ile	Ala	Met	Arg	Asp	Gln	Asn	Ala	Leu	Gln	His	Cys	Tyr
			100					105					110		
Trp	Lys	Tyr	Leu	Ile	Val	Asp	Glu	Gly	His	Arg	Ile	Lys	Asn	Met	Lys
		115					120					125			
Cys	Arg	Leu	Ile	Arg	Glu	Leu	Lys	Arg	Phe	Asn	Ala	Asp	Asn	Lys	Leu
	130					135					140				
Leu	Leu	Thr	Gly	Thr	Pro	Leu	Gln	Asn	Asn	Leu	Ser	Glu	Leu	Trp	Ser
145					150					155					160
Leu	Leu	Asn	Phe	Leu	Leu	Pro	Asp	Val	Phe	Asp	Asp	Leu	Lys	Ser	Phe
			165						170					175	
Glu	Ser	Trp	Phe	Asp	Ile	Thr	Ser	Leu	Ser	Glu	Thr	Ala	Glu	Asp	Ile
			180					185					190		
Ile	Ala	Lys	Glu	Arg	Glu	Gln	Asn	Val	Leu	His	Met	Leu	His	Gln	Ile
		195					200					205			
Leu	Thr	Pro	Phe	Leu	Leu	Arg	Leu	Lys	Ser	Asp	Val	Ala	Leu	Glu	
	210					215				220					
Val	Pro	Pro	Lys	Arg	Glu	Val	Val	Val	Tyr	Ala	Pro	Leu	Cys	Asn	Lys
225					230					235					240
Gln	Glu	Ile	Phe	Tyr	Thr	Ala	Ile	Val	Asn	Arg	Thr	Ile	Ala	Asn	Met
			245						250					255	
Phe	Gly	Ser	Cys	Glu	Lys	Glu	Thr	Val	Glu	Leu	Ser	Pro	Thr	Gly	Arg
			260					265					270		
Pro	Lys	Arg	Arg	Ser	Arg	Lys	Ser	Ile	Asn	Tyr	Ser	Glu	Leu	Asp	Gln
		275					280					285			
Phe	Pro	Ser	Glu	Leu	Glu	Lys	Leu	Ile	Ser	Gln	Ile	Gln	Pro	Glu	Val
	290					295					300				
Asn	Arg	Glu	Arg	Thr	Val	Glu	Gly	Asn	Ile	Pro	Ile	Glu	Ser	Glu	
305					310				315						320
Val	Asn	Leu	Lys	Leu	Arg	Asn	Ile	Met	Met	Leu	Leu	Arg	Lys	Cys	Cys
			325						330					335	
Asn	His	Pro	Tyr	Met	Ile	Glu	Tyr	Pro	Ile	Asp	Pro	Val	Thr	Gln	Glu
			340					345					350		

- 42 -

Phe Lys Ile Asp Glu Glu Leu Val Thr Asn Ser Gly Lys Phe Leu Ile
 355 360 365
 Leu Asp Arg Met Leu Pro Glu Leu Lys Lys Arg Gly His Lys Val Leu
 370 375 380
 Val Phe Ser Gln Met Thr Ser Met Leu Asp Ile Leu Met Asp Tyr Cys
 385 390 395 400
 His Leu Arg Asn Phe Ile Phe Ser Arg Leu Asp Gly Ser Met Ser Tyr
 405 410 415
 Ser Glu Arg Glu Lys Asn Ile Tyr Ser Phe Asn Thr Asp Pro Asp Val
 420 425 430
 Phe Leu Phe Leu Val Ser Thr Arg Ala Gly Gly Leu Gly Ile Asn Leu
 435 440 445
 Thr Ala Ala Asp Thr Val Ile Ile Tyr Asp Ser Asp Trp Asn Pro Gln
 450 455 460
 Ser Asp Leu Gln Ala Gln Asp Arg Cys His Arg Ile Gly Gln Thr Lys
 465 470 475 480
 Pro Val Val Val Tyr Arg Leu Val Thr Ala Asn Thr Ile Asp Gln Lys
 485 490 495
 Ile Val Glu Arg Ala Ala Ala Lys Arg Lys Leu Glu Lys Leu Ile Ile
 500 505 510
 His Lys Asn His Phe Lys Gly Gly Gln Ser Gly Leu Ser Gln Ser Lys
 515 520 525
 Asn Phe Leu Asp Ala Lys Glu Leu Met Glu Leu Leu Lys Ser Arg Asp
 530 535 540
 Tyr Glu Arg Glu Val Lys Gly Ser Arg Glu Lys Val Ile Ser Asp Glu
 545 550 555 560
 Asp Leu Glu Leu Leu Leu Asp Arg Ser Asp Leu Ile Asp Gln Met Lys
 565 570 575
 Ala Ser Arg Pro Ile Lys Gly Lys Thr Gly Ile Phe Lys Ile Leu Glu
 580 585 590
 Asn Ser Glu Asp Ser Ser Ala Glu Cys Leu Phe
 595 600

<210> 5
 <211> 1052
 <212> PRT
 <213> Homo sapiens

<400> 5
 Met Ser Ser Ala Ala Glu Pro Pro Pro Pro Pro Pro Glu Ser Ala
 1 5 10 15
 Pro Ser Lys Pro Ala Ala Ser Ile Ala Ser Gly Gly Ser Asn Ser Ser
 20 25 30
 Asn Lys Gly Gly Pro Glu Gly Val Ala Ala Gln Ala Val Ala Ser Ala
 35 40 45
 Ala Ser Ala Gly Pro Ala Asp Ala Glu Met Glu Glu Ile Phe Asp Asp
 50 55 60
 Ala Ser Pro Gly Lys Gln Lys Glu Ile Gln Glu Pro Asp Pro Thr Tyr
 65 70 75 80
 Glu Glu Lys Met Gln Thr Asp Arg Ala Asn Arg Phe Glu Tyr Leu Leu
 85 90 95
 Lys Gln Thr Glu Leu Phe Ala His Phe Ile Gln Pro Ala Ala Gln Lys
 100 105 110
 Thr Pro Thr Ser Pro Leu Lys Met Lys Pro Gly Arg Pro Arg Ile Lys
 115 120 125
 Lys Asp Glu Lys Gln Asn Leu Leu Ser Val Gly Asp Tyr Arg His Arg
 130 135 140
 Arg Thr Glu Gln Glu Glu Asp Glu Glu Leu Leu Thr Glu Ser Ser Lys
 145 150 155 160
 Ala Thr Asn Val Cys Thr Arg Phe Glu Asp Ser Pro Ser Tyr Val Lys
 165 170 175
 Trp Gly Lys Leu Arg Asp Tyr Gln Val Arg Gly Leu Asn Trp Leu Ile

- 44 -

```

        675                680                685
Asn Glu Lys Leu Ser Lys Met Gly Glu Ser Ser Leu Arg Asn Phe Thr
 690                695                700
Met Asp Thr Glu Ser Ser Val Tyr Asn Phe Glu Gly Glu Asp Tyr Arg
705                710                715                720
Glu Lys Gln Lys Ile Ala Phe Thr Glu Trp Ile Glu Pro Pro Lys Arg
        725                730                735
Glu Arg Lys Ala Asn Tyr Ala Val Asp Ala Tyr Phe Arg Glu Ala Leu
        740                745                750
Arg Val Ser Glu Pro Lys Ala Pro Lys Ala Pro Arg Pro Pro Lys Gln
        755                760                765
Pro Asn Val Gln Asp Phe Gln Phe Phe Pro Pro Arg Leu Phe Glu Leu
        770                775                780
Leu Glu Lys Glu Ile Leu Phe Tyr Arg Lys Thr Ile Gly Tyr Lys Val
785                790                795                800
Pro Arg Asn Pro Glu Leu Pro Asn Ala Ala Gln Ala Gln Lys Glu Glu
        805                810                815
Gln Leu Lys Ile Asp Glu Ala Glu Ser Leu Asn Asp Glu Glu Leu Glu
        820                825                830
Glu Lys Glu Lys Leu Leu Thr Gln Gly Phe Thr Asn Trp Asn Lys Arg
        835                840                845
Asp Phe Asn Gln Phe Ile Lys Ala Asn Glu Lys Trp Gly Arg Asp Asp
        850                855                860
Ile Glu Asn Ile Ala Arg Glu Val Glu Gly Lys Thr Pro Glu Glu Val
865                870                875                880
Ile Glu Tyr Ser Ala Val Phe Trp Glu Arg Cys Asn Glu Leu Gln Asp
        885                890                895
Ile Glu Lys Ile Met Ala Gln Ile Glu Arg Gly Glu Ala Arg Ile Gln
        900                905                910
Arg Arg Ile Ser Ile Lys Lys Ala Leu Asp Thr Lys Ile Gly Arg Tyr
        915                920                925
Lys Ala Pro Phe His Gln Leu Arg Ile Ser Tyr Gly Thr Asn Lys Gly
        930                935                940
Lys Asn Tyr Thr Glu Glu Asp Arg Phe Leu Ile Cys Met Leu His
945                950                955                960
Lys Leu Gly Phe Asp Lys Glu Asn Val Tyr Asp Glu Leu Arg Gln Cys
        965                970                975
Ile Arg Asn Ser Pro Gln Phe Arg Phe Asp Trp Phe Leu Lys Ser Arg
        980                985                990
Thr Ala Met Glu Leu Gln Arg Arg Cys Asn Thr Leu Ile Thr Leu Ile
        995                1000                1005
Glu Arg Glu Asn Met Glu Leu Glu Glu Lys Glu Lys Ala Glu Lys Lys
1010                1015                1020
Lys Arg Gly Pro Lys Pro Ser Thr Gln Lys Arg Lys Met Asp Gly Ala
1025                1030                1035                1040
Pro Asp Gly Arg Gly Arg Lys Lys Lys Leu Lys Leu
        1045                1050

```

```

<210> 6
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<223> /note= "synthetic construct"

```

```

<400> 6
tcaaggagat gattcgggcg t

```

21

```

<210> 7
<211> 22
<212> DNA
<213> Artificial Sequence

```

- 45 -

```
<220>
<223> /note= "synthetic construct"

<400> 7
aaaggaccca ttacagaac ac                                     22

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<223> /note= "synthetic construct"

<400> 8
gctggaaggg aaagcttaac aacc                                     24

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<223> /note= "synthetic construct"

<400> 9
acactgccat cgattctgca aacc                                     24
```

- 46 -

We claim:

1. An isolated nucleic acid molecule comprising a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, said gene occupying a segment of
5 said chromosome 5, lower arm, flanked on the centromeric side within 20 kilobases by a gene encoding a zinc-finger protein and on the telomeric side within 1 kilobase by a gene encoding a glutamic acid tRNA, the disruption of
10 said gene being associated with DNA hypomethylation.
2. The nucleic acid molecule of claim 1, wherein said gene is composed of exons that form an open reading frame having a sequence that encodes a
15 polypeptide about 750-850 amino acids in length.
3. A cDNA molecule comprising the exons of the nucleic acid molecule of claim 2.
- 20 4. The nucleic acid molecule of claim 2, wherein said open reading frame encodes an amino acid sequence substantially the same as SEQ ID NO:2.
- 25 5. The nucleic acid molecule of claim 4, wherein said open reading frame encodes amino acid SEQ ID NO:2.
6. The nucleic acid molecule of claim 5, which comprises an open reading frame of SEQ ID NO:1.
30
7. A recombinant DNA molecule, comprising a vector having an insert that includes the nucleic acid molecule of claim 1.
- 35 8. The recombinant DNA molecule of claim 7,

- 47 -

which is cosmid C38, ATCC Accession No. 207208.

9. An oligonucleotide between about 10 and 100 nucleotides in length, which specifically hybridizes with
5 a portion of the nucleic acid molecule of claim 1.

10. An isolated nucleic acid molecule which is a gene, the disruption of which is associated with DNA hypomethylation, having a sequence selected from the
10 group consisting of:

a) SEQ ID NO:1;

b) an allelic variant or natural mutant of
SEQ ID NO:1;

c) a sequence hybridizing with part or
15 all of SEQ ID NO:1 or its complement and encoding a polypeptide substantially the same as part or all of a polypeptide encoded by SEQ ID NO:1;

d) a sequence encoding part or all of a polypeptide having amino acid SEQ ID NO:2; and

20 e) a sequence encoding part or all of a polypeptide contained in the cosmid clone C38, designated ATCC Accession No. 207208.

11. A polypeptide produced by expression of an
25 isolated nucleic acid molecule comprising part or all of an open reading frame of a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, said gene occupying a segment of said chromosome 5, lower arm, flanked on the centromeric side within 20 kilobases by a gene encoding a
30 zinc-finger protein and on the telomeric side within 1 kilobase by a gene encoding a glutamic acid tRNA, the disruption of said gene being associated with DNA hypomethylation.

35 12. The polypeptide of claim 11, produced by

- 48 -

expression of a sequence selected from the group consisting of:

- a) SEQ ID NO:1;
 - b) an allelic variant or natural mutant of
5 SEQ ID NO:1;
 - c) a sequence hybridizing with part or all of SEQ ID NO:1 or its complement and encoding a polypeptide substantially the same as part or all of a polypeptide encoded by SEQ ID NO:1;
 - 10 d) a sequence encoding part or all of a polypeptide having amino acid SEQ ID NO:2; and
 - e) a sequence encoding part or all of a polypeptide contained in the clone designated ATCC Accession No. 207208.
- 15
13. The polypeptide of claim 11, having the amino acid sequence of part or all of SEQ ID NO:2.
14. An antibody immunologically specific for
20 the polypeptide of claim 11.
15. An isolated nucleic acid molecule having a sequence substantially the same as SEQ ID NO:3.
- 25
16. An isolated protein encoded by an *Arabidopsis thaliana* gene, said protein being a member of an SWI2/SNF2 family of polypeptides, loss of function of said protein being associated with DNA hypomethylation.
- 30
17. The protein of claim 16, encoded by a gene located on *A. thaliana* chromosome 5, lower arm, centromerically flanked within 20 kilobases by a zinc-finger-encoding gene and telomerically within one kilobase by a gene encoding a glutamic acid tRNA.
- 35

- 49 -

18. The protein of claim 16, encoded by a DNA .
segment on a recombinant cosmid C38, having ATCC
Accession No. 207208.

5

19. The protein of claim 16, having amino acid
SEQ ID NO:2.

20. A transgenic organism comprising the
10 nucleic acid molecule of claim 1.

21. The transgenic organism of claim 20, which
is a plant.

15 22. A method of stabilizing fidelity of DNA
methylation in an organism, comprising transforming the
organism with the nucleic acid molecule of claim 1.

20 23. A method of reducing or eliminating gene
silencing in a plant, comprising inhibiting or preventing
expression of an endogenous *DDM1* gene of the plant.

25 24. A method of introducing inbreeding
depression in a plant, comprising inhibiting or
preventing expression of an endogenous *DDM1* gene of the
plant.

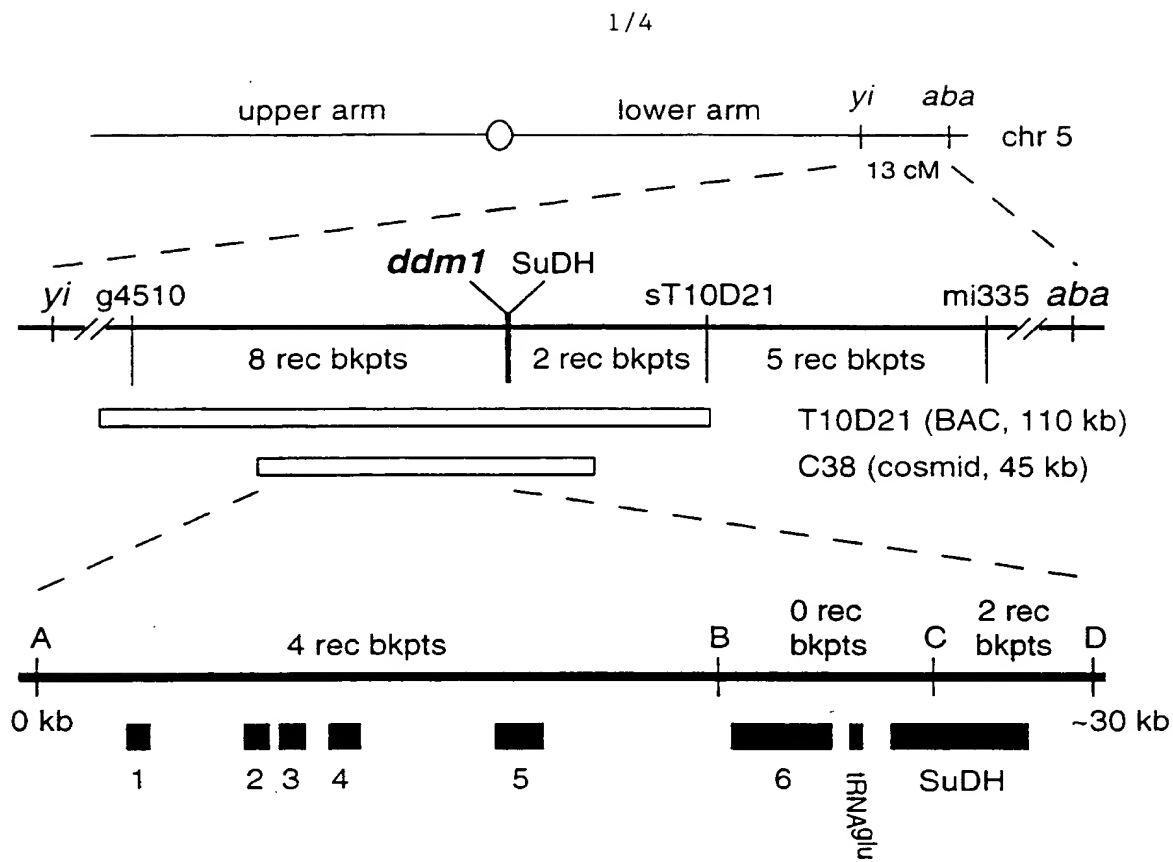


Figure 1

2/4

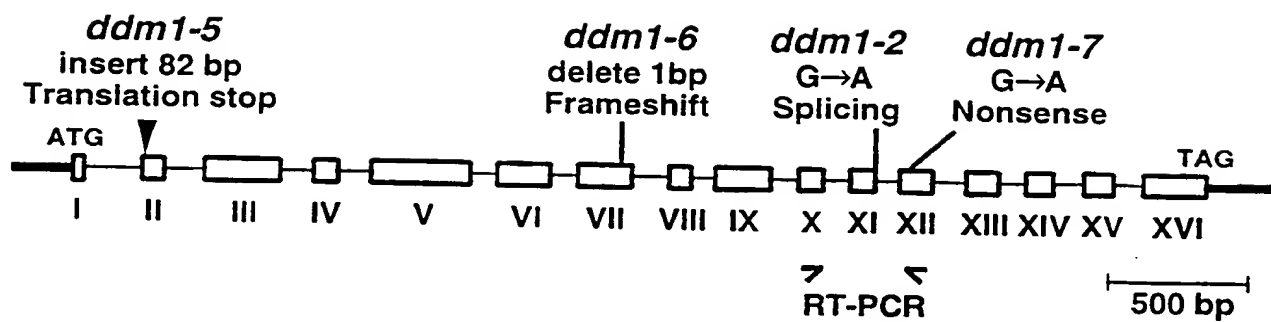


Fig. 2A

3/4

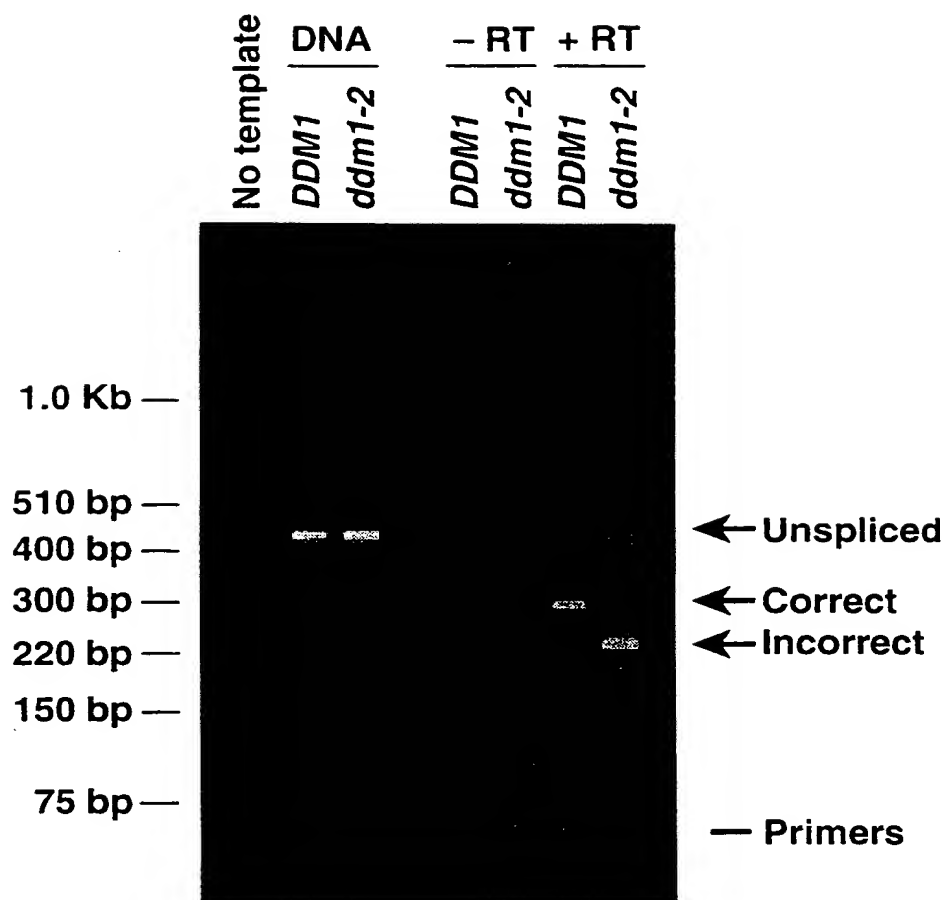


Fig. 2B

Al DDM1 Hs SNF2h	1	MVLSRKRKVI PASBMVSDGKTBDASG	SPTSVLN	EEENCERKSVTVVEE	ILLAN	NDSSLI	SEK	QORREQLLKL	DEEKANNA	GSAAVAPNL
	1	-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Hs SNF2h	96	RETOPKLDRL	QOTQLYSE	LLKME	DTINCH	SESE	SOKA	--EP	KTGRG	KKKAA
	74	QEPDPT	YEEKH	QOTDRANR	PEY	ELK	OTE	-YPA	HP	QPA
		-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Mm LSH Hs SNF2h	189	LQNELCP	-LLT	QGL	LSY	QGL	LSY	QGL	LSY	QGL
	1	-----	-----	-----	-----	-----	-----	-----	-----	-----
	166	TRPED	SPSY	VWGR	GRD	YOV	XG	GN	FL	IS
		-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Mm LSH Hs SNF2h	282	HGDKN	ORR	-K	NP	K	T	V	OP	K
	68	RGTE	DR	AK	L	V	K	N	H	K
	261	IGDK	EO	RA	P	V	D	L	L	P
		-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Mm LSH Hs SNF2h	378	LNE	LP	DF	T	S	H	D	F	E
	162	LNE	LP	DF	T	S	H	D	F	E
	351	LNE	LP	DF	T	S	H	D	F	E
		-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Mm LSH Hs SNF2h	471	HLG	-----	-----	-----	-----	-----	-----	-----	-----
	256	KPG	SC	ER	KT	VEL	S	PT	G	N
	435	ENB	-----	-----	-----	-----	-----	-----	-----	-----
		-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Mm LSH Hs SNF2h	511	YLV	PP	VE	EV	Q	CG	K	F	E
	351	QEK	-DE	EL	VT	NS	G	K	F	E
	470	PPT	TD	HL	VT	NS	G	K	F	E
		-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Mm LSH Hs SNF2h	606	GINT	A	A	D	T	C	I	L	Y
	445	GINT	A	A	D	T	C	I	L	Y
	565	GINT	A	A	D	T	C	I	L	Y
		-----	-----	-----	-----	-----	-----	-----	-----	-----
Al DDM1 Mm LSH Hs SNF2h	698	LKE	-D	E	T	A	E	D	R	L
	540	LKE	-D	E	T	A	E	D	R	L
	654	LKE	-D	E	T	A	E	D	R	L
		-----	-----	-----	-----	-----	-----	-----	-----	-----

Figure 3

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/09268

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/82 C07K14/415 C07K16/16 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JEDDELOH, J.A., ET AL. : "the DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis" GENES AND DEVELOPMENT, vol. 12, no. 11, 1 June 1998 (1998-06-01), pages 1714-1725, XP002114097 the whole document	23,24
X	MITTELSTEN-SCHEID, O., ET AL. : "release of epigenetic gene silencing by trans-acting mutations in Arabidopsis" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 95, January 1998 (1998-01), pages 632-637, XP002114098 cited in the application the whole document	23,24

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

2 September 1999

Date of mailing of the international search report

15/09/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Holtorf, S

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/09268

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KAKUTANI, T., ET AL. : "developmental abnormalities and epimutations associated with DNA hypomethylation mutations" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 93, October 1996 (1996-10), pages 12406-12411, XP002114099 page 12407, left column; page 12409, left column; Fig. 3 ---	1-6,10
Y	KAKUTANI, T., ET AL.: "characterization of an Arabidopsis thaliana hypomethylation mutant" NUCLEIC ACID RESEARCH, vol. 23, no. 1, 1995, pages 130-137, XP002049118 cited in the application abstract, last paragraph ---	1-6,10
A	KAKUTANI, T.: "genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana" THE PLANT JOURNAL, vol. 12, no. 6, 1997, pages 1447-1451, XP002114100 abstract, page 1448, right column ---	1-24
A	ROUNSLEY, S.D., ET AL. : "a BAC end sequence database for identifying minimal overlaps in Arabidopsis genomic sequencing . Update 4." EMBL SEQUENCE DATA LIBRARY, 29 May 1998 (1998-05-29), XP002114101 heidelberg, germany accession no. AQ010627 ---	1-24
A	VONGS, A., ET AL. : "Arabidopsis thaliana DNA methylation mutants" SCIENCE, vol. 260, June 1993 (1993-06), pages 1926-1928, XP002049119 cited in the application the whole document ---	1-24
A	WO 98 04725 A (UNIV YALE) 5 February 1998 (1998-02-05) abstract, page 10-14; examples 2 + 3, claims; --- -/--	1-24

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/09268

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PASZKOWSKI, J., ET AL.: "plant genes: the genetics of epigenetics" CURRENT BIOLOGY, vol. 8, no. 6, March 1998 (1998-03), pages R206-R208, XP002114102 the whole document ---	1-24
P,X	NAKAMURA, Y.: "structural analysis of Arabidopsis thaliana chromosome 5. IX. - unpublished" EMBL SEQUENCE DATA LIBRARY, 7 October 1998 (1998-10-07), XP002114103 heidelberg, germany accession no. AB018119 -----	1,2,10, 15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/09268

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9804725 A	05-02-1998	AU 4048097 A EP 0935666 A	20-02-1998 18-08-1999

Form PCT/ISA/210 (patent family annex) (July 1992)

THIS PAGE BLANK (USPTO)



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁶:C12N 15/82, C07K 14/415, 16/16, A01H
5/00

A1

(11) International Publication Number:

WO 99/55891

(43) International Publication Date:

4 November 1999 (04.11.99)

(21) International Application Number: PCT/US99/09268

(22) International Filing Date: 29 April 1999 (29.04.99)

(30) Priority Data:

60/083,612	30 April 1998 (30.04.98)	US
09/104,070	24 June 1998 (24.06.98)	US

(71) Applicant (for all designated States except US): WASHINGTON UNIVERSITY [US/US]: 600 South Euclid Avenue, St. Louis, MO 63110 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): RICHARDS, Eric, J. [US/US]: 4446 Westminster Place, St. Louis, MO 63108 (US). JEDDELOH, Jeffrey, A. [US/US]: 6756 W. Lakeridge Drive, New Market, MD 21774 (US).

(74) Agents: REED, Janet, E. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103 (US).

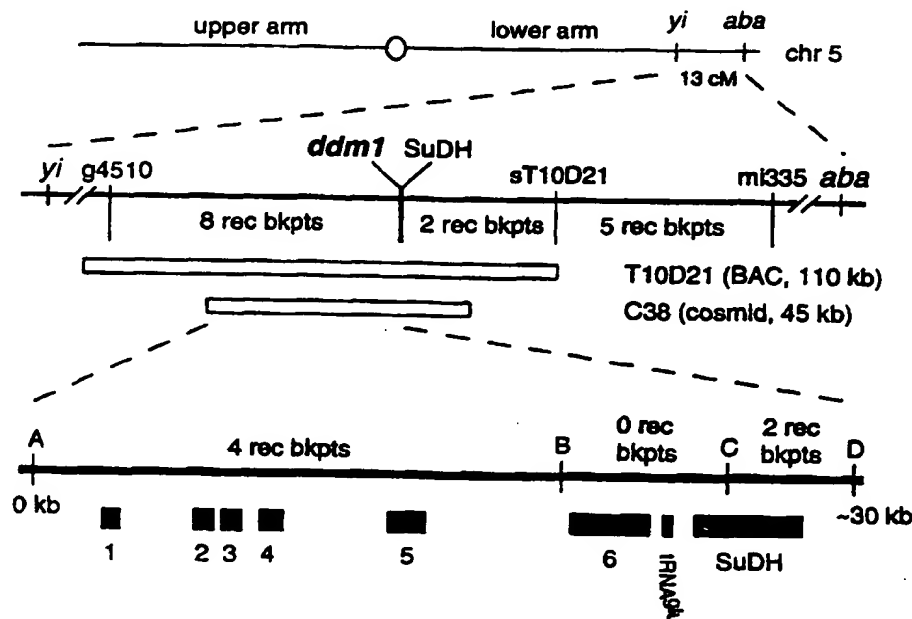
(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PLANT GENE THAT REGULATES DNA METHYLATION



(57) Abstract

A novel gene, *DDM1*, and its encoded protein are provided. The gene was isolated from a region of *Arabidopsis thaliana* chromosome 5. *DDM1* appears to be part of the SWI2/SNF2 family of chromatin-remodeling proteins. Disruption of the gene results in DNA hypomethylation, among other phenotypes. The *DDM1* gene defines a novel member of the DNA methylation system. Methods of using *DDM1*, and transgenic organisms comprising *DDM1*, are also provided.

*(Referred to in PCT Gazette No. 8/2000, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		



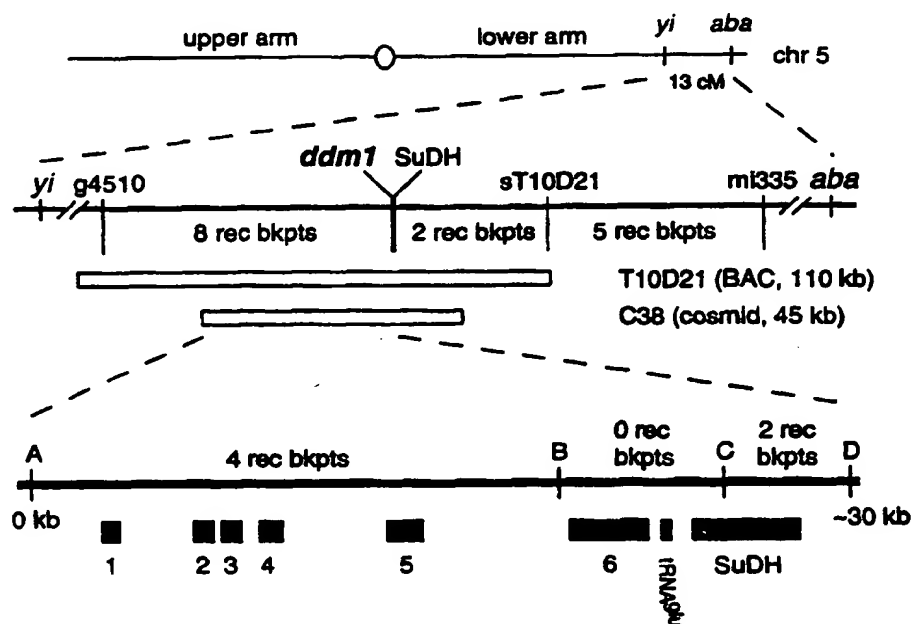
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/82, C07K 14/415, 16/16, A01H 5/00		A1	(11) International Publication Number: WO 99/55891
			(43) International Publication Date: 4 November 1999 (04.11.99)
(21) International Application Number: PCT/US99/09268		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 29 April 1999 (29.04.99)		<p>Published</p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
(30) Priority Data:			
60/083,612 30 April 1998 (30.04.98) US			
09/104,070 24 June 1998 (24.06.98) US			
(71) Applicant (for all designated States except US): WASHINGTON UNIVERSITY [US/US]; 600 South Euclid Avenue, St. Louis, MO 63110 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): RICHARDS, Eric, J. [US/US]; 4446 Westminster Place, St. Louis, MO 63108 (US). JEDDELOH, Jeffrey, A. [US/US]; 6756 W. Lakeridge Drive, New Market, MD 21774 (US).			
(74) Agents: REED, Janet, E. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103 (US).			

(54) Title: PLANT GENE THAT REGULATES DNA METHYLATION



(57) Abstract

A novel gene, *DDM1*, and its encoded protein are provided. The gene was isolated from a region of *Arabidopsis thaliana* chromosome 5. *DDM1* appears to be part of the SWI2/SNF2 family of chromatin-remodeling proteins. Disruption of the gene results in DNA hypomethylation, among other phenotypes. The *DDM1* gene defines a novel member of the DNA methylation system. Methods of using *DDM1*, and transgenic organisms comprising *DDM1*, are also provided.

*(Referred to in PCT Gazette No. 8/2000, Section II) ** (Referred to in PCT Gazette No. 14/2000, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

PLANT GENE THAT REGULATES DNA METHYLATION

Pursuant to 35 U.S.C. §202(c), it is acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the National Science Foundation, Grant
5 Nos. MCB9306266 and BIR9256779.

This application claims priority to U.S. Provisional Application Serial No. 60/_____, filed April 30, 1998, and to U.S. Application No. 09/104,070, filed June 24, 1998 the entireties of which are
10 incorporated by reference herein.

FIELD OF THE INVENTION

This invention relates to the field of plant molecular biology, genetic engineering and regulation of
15 gene expression. In particular, this invention provides a novel gene, *DDM1*, which plays an important role in the regulation of DNA methylation, and resultant regulation of gene expression, in plant genomic DNA.

20 BACKGROUND OF THE INVENTION

Various publications or patents are cited in this application to describe the state of the art to which the invention pertains. Each of these publications or patents is incorporated by reference herein.

25 Plant genomes contain substantial amounts of 5-methylcytosine. Up to 20-30% of the cytosines are methylated in the nuclear genome of many flowering plants. As in other organisms, methylation of cytosine

- 2 -

residues in plants occurs post-replicatively through the action of cytosine-DNA methyltransferases. Plant DNA methyltransferases have been characterized biochemically, and plant genes encoding these enzymes have been isolated
5 by virtue of their similarity to their mammalian counterparts.

Investigations of native plant genes and transgenic plants containing foreign genes have found a general correlation between transcriptional inactivity
10 and increased DNA methylation, consistent with evidence from mammalian systems. This evidence supports a role for cytosine methylation in maintaining transcriptional states.

The plant's need for developmental plasticity
15 and environmental interaction suggests that plants extensively employ epigenetic regulatory strategies. Such strategies rely on heritable, often reversible, changes in access to the underlying genetic information, but not alteration of the primary nucleotide sequence.
20 As one example, the alteration of DNA methylation is expected to perturb plant development significantly, provided that differential DNA methylation is an important component of epigenetic regulation in plants.

One paradigm linking DNA methylation and
25 developmental regulation comes from work on the mouse, where average genome cytosine methylation levels in embryonic lineages drop sharply in the early cleavages following fertilization, then rise again around the time of implantation. In plants, a similar pattern has been
30 observed in studies of DNA methylation content in pollen and post-embryonic tissue of varying age. Information from such studies indicates that there is a gradual rise in 5-methylcytosine levels in post-embryonic tissues

- 3 -

produced by meristems at positions further from the base of the plant (i.e., tissues of increasing age). Genetic studies of transposon systems in maize also demonstrate an age-dependent gradient of increasing epigenetic
5 modification, which is correlated with DNA methylation.

Both biochemical and genetic approaches have been taken to alter DNA methylation in eucaryotic organisms. Methylation inhibitor treatments have induced developmental abnormalities in many plant species.
10 Transgenic plants expressing antisense molecules specific for a native cytosine methyltransferase gene have been found to exhibit genomic hypomethylation, presumably due to the antisense interference with expression of the gene.

15 In another approach, mutants of *Arabidopsis thaliana* have been isolated, which show a decrease in DNA methylation (*ddm*) resulting in reduced nuclear 5-methylcytosine levels. The best characterized mutations define the *DDM1* gene. Homozygotes carrying recessive
20 *ddm1* alleles contain 30% of the wild-type levels of 5-methylcytosine. The *ddm1* mutations do not map to the two known cytosine-DNA methyltransferase genes of *A. thaliana*, nor do they affect DNA methyltransferase activity detectable in nuclear extracts (Kakutani et al.,
25 *Nuc. Acids Res.* **23**: 130-137, 1995). In addition, *ddm1* mutations do not appear to affect the metabolism of the active methyl group donor, S-adenosylmethionine (Kakutani et al., 1995, *supra*).

For the foregoing reasons, the *DDM1* gene
30 product is likely to be a novel component of the DNA methylation system, or involved in determining the cellular context (e.g., chromatin structure, subnuclear localization) of the methylation reaction. Consequently,

- 4 -

it would be a clear advance in the art of plant molecular and cellular biology to identify and isolate the *DDM1* gene and/or its encoded protein. Such a gene and protein would find utility for the purpose of modifying the methylation status of a selected genome and thereby altering one or more regulatory features of gene expression from that genome.

SUMMARY OF THE INVENTION

10 A novel gene, *DDM1*, and its encoded protein are provided in accordance with the present invention. The gene has been identified as a novel element of the DNA methylation system.

15 In one aspect of the invention, an isolated nucleic acid molecule comprising a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, is provided. The gene occupies a segment of chromosome 5, lower arm, which is flanked on the centromeric side within 20 kilobases by a gene encoding a zinc-finger protein and on the telomeric side within 1 kilobase by a gene encoding a glutamic acid tRNA. Disruption of the gene is associated with DNA hypomethylation. The gene encodes a polypeptide of about 764 amino acids in length. The nucleotide sequence of the *DDM1* gene is set forth
20 herein as SEQ ID NO:1 and its deduced amino acid sequence as SEQ ID NO:2. In SEQ ID NO:1, the regions of the gene that comprise coding sequence are indicated.

25 In another aspect of the invention, an isolated *DDM1* gene is provided, having a sequence selected from the group consisting of: (a) SEQ ID NO:1; (b) an allelic variant or natural mutant of SEQ ID NO:1; (c) a sequence hybridizing with part or all of SEQ ID NO:1 or its complement and encoding a polypeptide substantially the

30

- 5 -

same as part or all of a polypeptide encoded by SEQ ID NO:1; (d) a sequence encoding part or all of a polypeptide having amino acid SEQ ID NO:2; and (e) a sequence encoding part or all of a polypeptide contained
5 in the cosmid clone C38, designated ATCC Accession No. 207208.

According to another aspect of the invention, a polypeptide is provided, which is produced by expression of an isolated nucleic acid molecule comprising part or
10 all of an open reading frame of a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, the gene occupying a segment of chromosome 5, lower arm, flanked on the centromeric side within 20 kilobases by a gene encoding a zinc-finger protein and on the telomeric side
15 within 1 kilobase by a gene encoding a glutamic acid tRNA. This polypeptide preferably has the amino acid sequence of part or all of SEQ ID NO:2.

According to another aspect of the invention, an isolated protein encoded by an *Arabidopsis thaliana* gene is provided, which is a member of an SWI2/SNF2
20 family of polypeptides. Loss of function of the protein is associated with DNA hypomethylation. The protein is encoded by a gene located on *A. thaliana* chromosome 5, lower arm, centromerically flanked within 20 kilobases by
25 a zinc finger-encoding gene and telomerically within one kilobase by a gene encoding a glutamic acid tRNA.

According to another aspect of the invention, a transgenic organism comprising the *DDM1* gene is provided. In one embodiment, the transgenic organism is a plant.

30 In other aspects of the invention, methods are provided for stabilizing fidelity of DNA methylation in an organism, which comprise transforming the organism with the *DDM1* gene. Methods are also provided for

- 6 -

reducing or eliminating gene silencing in a plant, or for inducing inbreeding depression in a plant, which comprise inhibiting or preventing expression of an endogenous *DDM1* gene of the plant.

5 These aspects of the invention, as well as other features and advantages of the invention, will be described in greater detail in the description and examples set forth below.

10 BRIEF DESCRIPTION OF THE DRAWINGS

 Figure 1. Map-based isolation of the *A. thaliana* *DDM1* gene. A genetic map of the region of *A. thaliana* chromosome 5 containing the *DDM1* gene is shown at the top of the figure (see Example 1). The relative
15 sizes of the genetic intervals were determined by the number of recombination breakpoints (rec bkpts) scored in a panel of recombinant lines containing cross-overs between flanking markers *yi* and *aba*. The regions represented in genomic clones T10D21 and C38 are denoted
20 by the open boxes below the genetic map. The ~30 kb interval containing the *DDM1* gene, defined by the genetic markers A and D, is shown at the bottom of the figure. The number of recombination breakpoints scored between markers A - D and *ddm1-2* are indicated. The position of
25 predicted coding regions in the interval are numbered and shown below the physical map. BAC, bacterial artificial chromosome; SuDH, succinate dehydrogenase structural gene.

 Figure 2. *DDM1* gene structure and
30 identification. Fig. 2A: The intron/exon structure of the *DDM1* gene. Protein-coding exons are shown as open boxes, with the start and stop codons indicated. Introns are depicted as thin lines. The position and nature of

- 7 -

four *ddm1* alleles are indicated above the exon/intron map. **Fig. 2B:** RT-PCR analysis of *ddm1-2* and wild-type *DDM1* transcripts. The approximate positions of oligonucleotide primers used in the analysis are shown below the map in Fig. 2A. Amplifications were done on either genomic templates (DNA), first-strand cDNA templates (+RT, plus reverse transcriptase), or mock-synthesized cDNA (-RT, minus reverse transcriptase). Amplified products were separated on a 3% agarose gel and visualized after ethidium bromide staining. Amplification from cDNA representing the properly spliced transcript resulted in a ~280 bp product. The nucleotide sequence of the ~220 bp product amplified from *ddm1-2* cDNA template indicated that the mutation leads to use of an alternate splice donor 56 bp upstream of the wild-type splice donor site.

Figure 3. The *A. thaliana DDM1* gene encodes a SWI2/SNF2-like protein. The deduced primary amino acid sequence of *DDM1* (At *DDM1*) is aligned with two other SWI2/SNF2-like protein sequences, *Mus musculus* lymphocyte specific helicase (Mm LSH; SEQ ID NO:4) and human SNF2h (Hs SNF2h; SEQ ID NO:5). Sequence identities are indicated by black boxes and conservative changes are shaded. The positions of the eight signature motifs characteristic of SNF2 family proteins are indicated below the aligned sequences. Amino acid coordinates are indicated on the left; only the N-terminal 730 amino acids (of 1052 total) are shown for human SNF2h, though SEQ ID NO:5 shows the entire protein sequence. The deletion/frameshift caused by the *ddm1-2* allele occurs at amino acid 524. The *ddm1-6* frameshift occurs at amino acid 379, leading to translation of an additional 52 amino acids out of frame. The *ddm1-7* nonsense mutation

- 8 -

occurs at amino acid 549. Dashes indicate gaps introduced by the CLUSTAL W algorithm to maximize alignment (Thompson et al., Nucleic Acids Res. 22: 4673-4680, 1994). The alignment was processed by BOXSHADE v. 3.21.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Various terms relating to the biological molecules of the present invention are used throughout the specification and claims.

With reference to nucleic acids of the invention, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule.

With respect to RNA molecules of the invention the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

With respect to protein, the term "isolated protein" or "isolated and purified protein" is sometimes

- 9 -

used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein which has been sufficiently separated
5 from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid,
10 oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods,
15 agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

Nucleic acid sequences and amino acid sequences can be compared using computer programs that align the similar sequences of the nucleic or amino acids thus
20 define the differences. In the comparisons made in the present invention, the CLUSTLW program and parameters employed therein were utilized (Thompson et al., 1994, *supra*). However, equivalent alignments and similarity/identity assessments can be obtained through
25 the use of any standard alignment software. For instance, the GCG Wisconsin Package version 9.1, available from the Genetics Computer Group in Madison, Wisconsin, and the default parameters used (gap creation penalty=12, gap extension penalty=4) by that program may
30 also be used to compare sequence identity and similarity.

The term "substantially the same" refers to nucleic acid or amino acid sequences having sequence variation that do not materially affect the nature of the

- 10 -

protein (i.e. the structure, stability characteristics, substrate specificity and/or biological activity of the protein). With particular reference to nucleic acid sequences, the term "substantially the same" is intended to refer to the coding region and to conserved sequences governing expression, and refers primarily to degenerate codons encoding the same amino acid, or alternate codons encoding conservative substitute amino acids in the encoded polypeptide. With reference to amino acid sequences, the term "substantially the same" refers generally to conservative substitutions and/or variations in regions of the polypeptide not involved in determination of structure or function.

The terms "percent identical" and "percent similar" are also used herein in comparisons among amino acid and nucleic acid sequences. When referring to amino acid sequences, "percent identical" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical amino acids in the compared amino acid sequence by a sequence analysis program. "Percent similar" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical or conserved amino acids. Conserved amino acids are those which differ in structure but are similar in physical properties such that the exchange of one for another would not appreciably change the tertiary structure of the resulting protein. Conservative substitutions are defined in Taylor (1986, J. Theor. Biol. 119:205). When referring to nucleic acid molecules, "percent identical" refers to the percent of the nucleotides of the subject nucleic acid sequence that have been matched to identical nucleotides by a sequence analysis program.

- 11 -

With respect to antibodies, the term "immunologically specific" refers to antibodies that bind to one or more epitopes of a protein of interest, but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

With respect to oligonucleotides or other single-stranded nucleic acid molecules, the term "specifically hybridizing" refers to the association between two single-stranded nucleic acid molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

A "coding sequence" or "coding region" refers to a nucleic acid molecule having sequence information necessary to produce a gene product, when the sequence is expressed.

The term "operably linked" or "operably inserted" means that the regulatory sequences necessary for expression of the coding sequence are placed in a nucleic acid molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement other transcription control elements (e.g. enhancers) in an expression vector.

Transcriptional and translational control sequences are DNA regulatory sequences, such as

- 12 -

promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell. In particular, as used herein, the term "DNA transcriptional response element" refers to a DNA sequence specifically recognized for binding by a DNA binding protein characterized as a transcriptional regulator (either activator or suppressor).

The terms "promoter", "promoter region" or "promoter sequence" refer generally to transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. Typically, a promoter is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The typical 5' promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A "vector" is a replicon, such as plasmid, phage, cosmid, or virus to which another nucleic acid segment may be operably inserted so as to bring about the replication or expression of the segment.

The term "nucleic acid construct" or "DNA construct" is sometimes used to refer to a coding sequence or sequences operably linked to appropriate regulatory sequences and inserted into a vector for transforming a cell. This term may be used interchangeably with the term "transforming DNA". Such a

- 13 -

nucleic acid construct may contain a coding sequence for a gene product of interest, along with a selectable marker gene and/or a reporter gene.

5 The term "reporter gene" refers to genetic sequences which may be operably linked to a promoter region forming a transgene, such that expression of the reporter gene coding region is regulated by the promoter and expression of the transgene is readily assayed.

10 The term "selectable marker gene" refers to a gene product that when expressed confers a selectable phenotype, such as antibiotic resistance, on a transformed cell or plant.

The term "DNA construct" is sometimes used herein to refer to genetic sequence used to transform
15 plants and generate progeny transgenic plants. These constructs may be administered to plants in a viral or plasmid vector. Other methods of delivery such as Agrobacterium T-DNA mediated transformation and transformation using the biolistic process are also
20 contemplated to be within the scope of the present invention. The transforming DNA may be prepared according to standard protocols such as those set forth in "Current Protocols in Molecular Biology", eds. Frederick M. Ausubel et al., John Wiley & Sons, 1995.

25 A cell has been "transformed" or "transfected" by exogenous or heterologous DNA construct when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and plant
30 cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells

- 14 -

through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A
5 "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

10

II. Description of *DDM1* and its Encoded Polypeptide

In accordance with the present invention, a
15 novel gene, *DDM1*, has been isolated from the flowering plant *Arabidopsis thaliana*. Through analysis of mutant plants, this gene has been identified as important for the maintenance of proper genomic cytosine methylation, and its function appears to be necessary to maintain gene
20 silencing. Biochemical and molecular genetic results indicate that *DDM1* encodes a novel component of the DNA methylation machinery.

We have isolated the *DDM1* gene from *A. thaliana* using a map-based cloning approach, which is described in
25 detail in Example 1 and shown in Figure 1. Briefly, the *DDM1* gene was initially localized to the bottom of the lower arm of chromosome 5 by reference to molecular markers segregating in an F2 family (parental cross: Columbia *ddm1/ddm1* X Landsberg erecta *DDM1/DDM1*). Next,
30 recombination breakpoints in the region surrounding a *ddm1* mutation were isolated by collecting cross-over chromosomes by reference to flanking genetic markers. The recombination breakpoints delimited a region of approximately 30 kilobases. Cloned DNA corresponding to
35 this genomic region was isolated by subcloning DNA from a

- 15 -

bacterial artificial chromosome (BAC) containing molecular markers mapping both proximal and distal to the *ddm1* marker. The nucleotide sequence of a single cosmid subclone encompassing the 30 kb region was determined to
5 identify six candidate genes, in addition to a tRNA gene and a previously identified succinate dehydrogenase structural gene.

The search for the *DDM1* gene focused on predicted genes 5 and 6, which fell in the center of the
10 genetic interval defined by recombination breakpoints with the *ddm1-2* marker. The *DDM1* gene was identified as predicted gene 6 based on DNA sequence alterations in four *ddm1* alleles (Figure 2). The EMS-generated *ddm1-2* mutation is a G to A transition in the splice donor site
15 of intron 11 that forces the use of an alternate splice donor site 56 bp upstream in exon 11 (Fig. 2B). The splicing defect leads to a deletion, a frameshift and premature translation termination upstream of predicted functional domains. The fast neutron-generated *ddm1-5*
20 (previously named *som8*; Mittelsten Scheid, O., Afsar, K. & Paszkowski, J. *Proc. Natl. Acad. Sci. USA* 95: 632-637, 1998).) allele contains an 82 bp insertion (1 bp deleted and replaced with 83 bp) in the second protein-coding exon, leading to an in-frame stop after 30 codons (15
25 wild-type codons plus 15 codons from the insertion). Premature translation termination is also predicted to result from two additional fast neutron alleles: *ddm1-6* (*som4*) corresponds to a frameshift (1 bp deletion) in exon 7 and *ddm1-7* (*som5*) is a nonsense mutation in exon
30 12. All four characterized *ddm1* alleles are expected to destroy or severely reduce gene function.

The wild-type *DDM1* gene encodes a predicted protein of 764 amino acids with a high degree of

- 16 -

similarity to SWI2/SNF2-like proteins. Members of the SWI2/SNF2 family are involved in various functions, including transcriptional co-activation, transcriptional co-repression, chromatin assembly and DNA repair.

5 Underlying these apparently diverse activities is the modification or disruption of protein-DNA interactions by multi-protein complexes which contain SWI2/SNF2-like components. Figure 3 shows an alignment among the deduced amino acid sequences of *A. thaliana* DDM1 and two
10 mammalian members of the SNF2 family, human SNF2h (SEQ ID NO:4; Arihara, T. et al., *Cytogenet. Cell Genet.* **81**, 191-193, 1998) and murine LSH (SEQ ID NO:5; lymphocyte specific helicase, LSH; Jarvis, C.D. et al. *Gene* **169**, 203-207, 1996). DDM1 contains the eight sequence motifs
15 diagnostic of SWI2/SNF2 family members (Bork, P. & Koonin, E.V. *Nucleic Acids Res.* **21**, 751-752, 1993). *A. thaliana* DDM1 and human SNF2h share 45 percent identity over the approximately 470 amino acid region comprising the signature motifs. Over a similar region, *A. thaliana*
20 DDM1 and murine LSH display approximately 50 percent identity, omitting the 47 residues (amino acids 276-322) apparently unique to LSH. Initial molecular phylogenetic analysis placed DDM1 in a small subfamily, within the SNF2 family, which contains proteins of unknown function, including murine LSH (Eisen, J.A. et al. *Nucleic Acids*
25 *Res.* **23**, 2715-2723, 1995). The proteins of known function most closely related to DDM1 are involved in chromatin remodeling and are grouped in the SNF2L/ISWI subfamily (Eisen et al., 1995, *supra*).

30 Without intending to be bound by any particular mechanism for the functionality of the *DDM1* gene product, analysis of the foregoing data indicates that the *DDM1* protein functions in the DNA methylation system by

- 17 -

affecting chromatin structure. Two general models for the *DDM1* action are envisioned. The *DDM1* protein may function as a transcriptional co-activator, similar to many SWI2/SNF2-like proteins, to increase the expression of a component of the DNA methylation system. *DDM1* does not affect DNA methyltransferase expression directly because *ddm1* mutant extracts contain wild-type methyltransferase activity (Kakutani et al., 1995, *supra*). However, an unidentified positive effector of DNA methylation may be a target. Alternatively, wild-type *DDM1* function may change chromatin structure to direct certain sequences to the methylation machinery or to facilitate the methylation of genomic substrates. The recently discovered interplay between cytosine methylation and histone acetylation, and the association of SWI2/SNF2-like proteins and histone deacetylases in chromatin remodeling complexes, makes it plausible that *DDM1* affects DNA methylation through modulation of histone modification or another aspect of chromatin structure. Another possibility is that *DDM1* plays a more direct role as a part of a nucleosome remodeling complex that increases the accessibility of the DNA methyltransferase to the hemimethylated substrates in newly replicated chromatin. The latter model is particularly attractive because it predicts that *ddm1* mutations will preferentially hypomethylate genomic sequences packaged in highly condensed chromatin while causing slower loss of methylation in more accessible sequences, consistent with the observed hypomethylation specificity of *ddm1* mutations. The isolation of the *Arabidopsis DDM1* gene in accordance with the present invention points to the importance of chromatin dynamics in the maintenance of cytosine methylation patterns and

- 18 -

identifies a novel component of the eukaryotic DNA methylation pathway.

A number of applications are contemplated for the novel gene of the invention and its encoded protein, and the discovery of the involvement of a *SWI2/SNF2*-like gene in the eucaryotic DNA methylation system. Such applications are described in greater detail below.

Although the *DDM1* genomic clone from *Arabidopsis thaliana* is described and exemplified herein, this invention is intended to encompass nucleic acid sequences and proteins from other organisms, including plants, yeast, insects and mammals, that are sufficiently similar to be used instead of the *Arabidopsis DDM1* nucleic acid and proteins for the purposes described below. These include, but are not limited to, allelic variants and natural mutants of SEQ ID NO:1, which are likely to be found in different species of plants or varieties of *Arabidopsis*. Because such variants are expected to possess certain differences in nucleotide and amino acid sequence, this invention provides an isolated *DDM1* nucleic acid molecule having at least about 60% (preferably 70% and more preferably over 80%) sequence homology in the coding regions with the nucleotide sequence set forth as SEQ ID NO:1 (and, most preferably, specifically comprising the coding region of SEQ ID NO:1). This invention also provides isolated polypeptide products of the open reading frames of SEQ ID NO:1, having at least about 60% (preferably 70% or 80% or greater) sequence homology with the amino acid sequences of SEQ ID NO:2. Because of the natural sequence variation likely to exist among *DDM1* genes, one skilled in the art would expect to find up to about 30-40% nucleotide sequence variation, while still maintaining

- 19 -

the unique properties of the *DDM1* gene and encoded polypeptide of the present invention. Such an expectation is due in part to the degeneracy of the genetic code, as well as to the known evolutionary success of conservative amino acid sequence variations, which do not appreciably alter the nature of the encoded protein. Accordingly, such variants are considered substantially the same as one another and are included within the scope of the present invention.

The following description sets forth the general procedures involved in practicing the present invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning procedures, such as those set forth in Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989) (hereinafter "Sambrook et al.") or Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1999) (hereinafter "Ausubel et al.") are used.

A. Preparation of *DDM1* Nucleic Acid Molecules, encoded Polypeptides and Antibodies Specific for the Polypeptides

1. Nucleic Acid Molecules

DDM1 nucleic acid molecules of the invention may be prepared by two general methods: (1) they may be synthesized from appropriate nucleotide triphosphates, or (2) they may be isolated from biological sources. Both methods utilize protocols well known in the art.

The availability of nucleotide sequence information, such as the cDNA having SEQ ID NO:1, enables preparation of an isolated nucleic acid molecule of the

- 20 -

invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramidite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. The resultant construct
5 may be purified according to methods known in the art, such as high performance liquid chromatography (HPLC). Long, double-stranded polynucleotides, such as a DNA molecule of the present invention, must be synthesized in stages, due to the size limitations inherent in current
10 oligonucleotide synthetic methods. Thus, for example, a long double-stranded molecule may be synthesized as several smaller segments of appropriate complementarity. Complementary segments thus produced may be annealed such that each segment possesses appropriate cohesive termini
15 for attachment of an adjacent segment. Adjacent segments may be ligated by annealing cohesive termini in the presence of DNA ligase to construct an entire long double-stranded molecule. A synthetic DNA molecule so constructed may then be cloned and amplified in an
20 appropriate vector.

DDM1 genes also may be isolated from appropriate biological sources using methods known in the art. In the exemplary embodiment of the invention, the *A. thaliana DDM1* clone was isolated from a BAC genomic
25 library of *A. thaliana*. In alternative embodiments, cDNA clones of *DDM1* may be isolated. A preferred means for isolating *DDM1* genes is PCR amplification using genomic templates and *DDM1*-specific primers.

In accordance with the present invention,
30 nucleic acids having the appropriate level sequence homology with part or all the coding regions of SEQ ID NO:1 may be identified by using hybridization and washing conditions of appropriate stringency. For example,

- 21 -

hybridizations may be performed, according to the method of Sambrook et al., using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 µg/ml denatured; fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 2X SSC and 0.1% SDS; (4) 2 hours at 45-55°C in 2X SSC and 0.1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology (Sambrook et al., 1989):

$$T_m = 81.5^{\circ}\text{C} + 16.6\text{Log} [\text{Na}^+] + 0.41(\% \text{ G+C}) - 0.63 (\% \text{ formamide}) - 600/\text{\#bp in duplex}$$

As an illustration of the above formula, using $[\text{Na}^+] = [0.368]$ and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C. Such a sequence would be considered substantially homologous to the sequences of the present invention.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a preferred embodiment, clones are maintained in plasmid cloning/expression vector, such as pGEM-T (Promega Biotech, Madison, WI) or pBluescript (Stratagene, La Jolla, CA), either of which is propagated in a suitable

E. coli host cell.

DDM1 nucleic acid molecules of the invention include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, this invention provides oligonucleotides (sense or antisense strands of DNA or RNA) having sequences capable of hybridizing with at least one sequence of a nucleic acid molecule of the present invention, such as selected segments of the DNA having SEQ ID NO:1. Such oligonucleotides are useful as probes for detecting *DDM1* genes or mRNA in test samples, e.g. by PCR amplification, or for the positive or negative regulation of expression of *DDM1* genes at or before translation of the mRNA into proteins.

The *DDM1* promoter and other expression regulatory sequences for *DDM1* are also expected to be useful in connection with the present invention. SEQ ID NO:1 shows about 550 bp of sequence upstream from the beginning of the coding region, which should contain such expression regulatory sequences. In addition, SEQ ID NO:3 constitutes about 5 kbp of additional upstream sequence, which should contain other regulatory sequences, such as enhancer elements.

25 2. Proteins

Polypeptides encoded by *DDM1* nucleic acids of the invention may be prepared in a variety of ways, according to known methods. If produced *in situ* the polypeptides may be purified from appropriate sources, e.g., plant parts.

Alternatively, the availability of nucleic acid molecules encoding the polypeptides enables production of the proteins using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into

- 23 -

an appropriate *in vitro* transcription vector, such a
pSP64 or pSP65 for *in vitro* transcription, followed by
cell-free translation in a suitable cell-free translation
system, such as wheat germ or rabbit reticulocytes. In
5 *in vitro* transcription and translation systems are
commercially available, e.g., from Promega Biotech,
Madison, Wisconsin or BRL, Rockville, Maryland.

According to a preferred embodiment, larger
quantities of DDM1-encoded polypeptide may be produced by
10 expression in a suitable procaryotic or eucaryotic
system. For example, part or all of a DNA molecule, such
as the coding portion of SEQ ID NO:1, may be inserted
into a plasmid vector adapted for expression in a
bacterial cell (such as *E. coli*) or a yeast cell (such as
15 *Saccharomyces cerevisiae*), or into a baculovirus vector
for expression in an insect cell. Such vectors comprise
the regulatory elements necessary for expression of the
DNA in the host cell, positioned in such a manner as to
permit expression of the DNA in the host cell. Such
20 regulatory elements required for expression include
promoter sequences, transcription initiation sequences
and, optionally, enhancer sequences.

The DDM1 polypeptide produced by gene
expression in a recombinant procaryotic or eucaryotic
25 system may be purified according to methods known in the
art. In a preferred embodiment, a commercially available
expression/secretion system can be used, whereby the
recombinant protein is expressed and thereafter secreted
from the host cell, to be easily purified from the
30 surrounding medium. If expression/secretion vectors are
not used, an alternative approach involves purifying the
recombinant protein by affinity separation, such as by
immunological interaction with antibodies that bind
specifically to the recombinant protein. Such methods
35 are commonly used by skilled practitioners.

The *DDM1*-encoded polypeptides of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures. Methods for analyzing the functional activity are available. For instance, DNA methylation levels are detectable by known methods. 5 Alternatively, the function of the *DDM1* gene product as part of a chromatin remodeling machine permits the use of *in vitro* assays for chromatin remodeling, which are known in the art (e.g., B.R. Cairns, *Trends in Biochem.* 23: 20-10 25, 1998).

The present invention also provides antibodies capable of immunospecifically binding to polypeptides of the invention. Polyclonal or monoclonal antibodies directed toward the polypeptide encoded by *DDM1* may be 15 prepared according to standard methods. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. In a preferred embodiment, antibodies are prepared, which react immunospecifically with various epitopes of the 20 *DDM1*-encoded polypeptides.

**B. Uses of *DDM1* Nucleic Acids,
Encoded Proteins and Antibodies**

1. *DDM1* Nucleic Acids

DDM1 nucleic acids may be used for a variety of 25 purposes in accordance with the present invention. The DNA, RNA, or fragments thereof may be used as probes to detect the presence of and/or expression of *DDM1* genes. Methods in which *DDM1* nucleic acids may be utilized as 30 probes for such assays include, but are not limited to: (1) *in situ* hybridization; (2) Southern hybridization (3) northern hybridization; and (4) assorted amplification reactions such as polymerase chain reactions (PCR).

The *DDM1* nucleic acids of the invention may 35 also be utilized as probes to identify related genes from

- 25 -

other species, including but not limited to, plants, yeast, insects and mammals, including humans. As is well known in the art and described above, hybridization stringencies may be adjusted to allow hybridization of nucleic acid probes with complementary sequences of varying degrees of homology. Thus, *DDM1* nucleic acids may be used to advantage to identify and characterize other genes of varying degrees of relation to the exemplary coding sequence of SEQ ID NO:1, thereby enabling further characterization of this family of genes. Additionally, they may be used to identify genes encoding proteins that interact with protein encoded by *DDM1* (e.g., by the "interaction trap" technique).

As discussed above and in greater detail in Example 1, the similarity among plant *DDM1* and its *SWI2/SNF2* counterparts in yeast, *Drosophila* and mammals indicates that the functional aspects of these proteins will also be conserved. Thus, *DDM1* is expected to play an important role in DNA methylation and resultant down-regulation of gene expression. Plants engineered to over-express *DDM1* can be expected to have improved fidelity of the DNA methylation system. The evidence suggests that loss of *DDM1* function leads to reduction in the efficiency of maintenance methylation due to reduced accessibility of the methyltransferase enzyme to the substrate. Hence, excess *DDM1* function could lead to an increase in the fidelity of the inheritance of DNA methylation thereby reducing the occurrence of spurious methylation mistakes which could compromise the organism's viability or fecundity. In fact, there are experimental data demonstrating that loss of *DDM1* function leads to stochastic hypermethylation, and epigenetic lesion formation, as well. For these reasons, *DDM1* overexpression lines are expected to have useful properties.

- 26 -

Transgenic plants expressing the *DDM1* gene or antisense nucleotides can be generated using standard plant transformation methods known to those skilled in the art. These include, but are not limited to,

5 *Agrobacterium* vectors, PEG treatment of protoplasts, biolistic DNA delivery, UV laser microbeam, gemini virus vectors, calcium phosphate treatment of protoplasts, electroporation of isolated protoplasts, agitation of cell suspensions with microbeads coated with the

10 transforming DNA, direct DNA uptake, liposome-mediated DNA uptake, and the like. Such methods have been published in the art. See, e.g., Methods for Plant Molecular Biology (Weissbach & Weissbach, eds., 1988); Methods in Plant Molecular Biology (Schuler & Zielinski,

15 eds., 1989); Plant Molecular Biology Manual (Gelvin, Schilperoort, Verma, eds., 1993); and Methods in Plant Molecular Biology - A Laboratory Manual (Maliga, Klessig, Cashmore, Gruissem & Varner, eds., 1994).

The method of transformation depends upon the

20 plant to be transformed. The biolistic DNA delivery method is useful for nuclear transformation. In another embodiment of the invention, *Agrobacterium* vectors are used to advantage for efficient transformation of plant nuclei.

25 In a preferred embodiment, the gene is introduced into plant nuclei in *Agrobacterium* binary vectors. Such vectors include, but are not limited to, BIN19 (Bevan, 1984) and derivatives thereof, the pBI vector series (Jefferson et al., 1987), and binary

30 vectors pGA482 and pGA492 (An, 1986).

The *DDM1* gene may be placed under a powerful constitutive promoter, such as the Cauliflower Mosaic Virus (CaMV) 35S promoter or the figwort mosaic virus 35S promoter. Transgenic plants expressing the *DDM1* gene

35 under an inducible promoter (either its own promoter or a

- 27 -

heterologous promoter) are also contemplated to be within the scope of the present invention. Inducible plant promoters include the tetracycline repressor/operator controlled promoter.

5 Using an *Agrobacterium* binary vector system for transformation, the *DDM1* coding region, under control of a constitutive or inducible promoter as described above, is linked to a nuclear drug resistance marker, such as kanamycin resistance. *Agrobacterium*-mediated
10 transformation of plant nuclei is accomplished according to the following procedure:

(1) the gene is inserted into the selected *Agrobacterium* binary vector;

(2) transformation is accomplished by co-
15 cultivation of plant tissue (e.g., leaf discs) with a suspension of recombinant *Agrobacterium*, followed by incubation (e.g., two days) on growth medium in the absence of the drug used as the selective medium (see, e.g., Horsch et al. 1985);

20 (3) plant tissue is then transferred onto the selective medium to identify transformed tissue; and

(4) identified transformants are regenerated to intact plants.

It should be recognized that the amount of
25 expression, as well as the tissue specificity of expression of the *DDM1* gene in transformed plants can vary depending on the position of their insertion into the nuclear genome. Such position effects are well known in the art. For this reason, several nuclear
30 transformants should be regenerated and tested for expression of the transgene.

In some instances, it may be desirable to down-regulate or inhibit expression of endogenous *DDM1* in plants possessing the gene. One clear benefit to
35 engineering a reduction of *DDM1* function is to reduce

- 28 -

gene (including transgene) silencing. Plant lines with reduced or absent DDM1 function are expected to be viable based on results obtained with *Arabidopsis*. Further, it has been shown that gene silencing is suppressed in *ddm1* *Arabidopsis* lines (Jeddeloh et al., *Genes Devel.* 12:1714-1725, 1998). There are two other beneficial characteristics of DDM1 deficient plant lines. First, alteration in DNA methylation leads to changes in flowering time, and as such, is a potentially powerful tool for manipulating plant development. (See, e.g., Richards, *Trends in Genetics* 13: 319-323, 1998), Second, *ddm1* mutant lines exhibit inbreeding depression (a reduction in vigor after inbreeding) (Richards, *Trends in Genetics*, 1998, *supra*), a characteristic which may be desirable to include in situations where proprietary germplasms in hybrid plants are at risk of unauthorized use. For instance, a genetically engineered hybrid (containing one or more useful transgenes) could be further engineered to down-regulate endogenous DDM1 expression. Unauthorized inbreeding of such lines would be discouraged because the progeny of such lines would lack vigor.

To achieve the aforementioned benefits associated with reduced gene expression, DDM1 nucleic acid molecules, or fragments thereof, may also be utilized to control the production of DDM1-encoded proteins. In one embodiment, full-length DDM1 antisense molecules or antisense oligonucleotides, targeted to specific regions of DDM1-encoded RNA that are critical for translation, are used. The use of antisense molecules to decrease expression levels of a pre-determined gene is known in the art. In a preferred embodiment, antisense molecules are provided *in situ* by transforming plant cells with a DNA construct which, upon

- 29 -

transcription, produces the antisense sequences. Such constructs can be designed to produce full-length or partial antisense sequences.

In another embodiment, overexpression of *DDM1* is induced to generate a co-suppression effect. This excess expression serves to promote down-regulation of both endogenous and exogenous *DDM1* genes.

Optionally, transgenic plants can be created containing mutations in the region encoding the active site of *DDM1*. This embodiment may be preferred in certain instances.

From the foregoing discussion, it can be seen that *DDM1* and its homologs will be useful for introducing alterations in gene expression in an organism, for a variety of purposes. As described above, for instance, the *Arabidopsis DDM1* gene can be used to isolate mutants or engineer organisms that express reduced function of *DDM1* orthologs. Based on results in *Arabidopsis*, such mutants or engineered organisms are expected to be viable and display valuable characteristics, such as inbreeding depression and a reduction in gene silencing. In addition, we anticipate that dysfunction in human *DDM1* orthologs may contribute to diseases that involve alterations in DNA methylation, including cancer (Baylin, S.B. et al., *Adv. Cancer Res.* 72: 141-196, 1998) and immunodeficiency/ chromosome instability/facial anomalies syndrome (ICF) (Smeets, D.F.C.M. et al., *Hum. Genet.* 94: 240-246, 1994).

2. DDM1 Proteins and Antibodies

Purified *DDM1*-encoded proteins, or fragments thereof, may be used to produce polyclonal or monoclonal antibodies which also may serve as sensitive detection reagents for the presence and accumulation of *DDM1*-encoded

- 30 -

protein in cultured cells or tissues and in intact organisms. Recombinant techniques enable expression of fusion proteins containing part or all of the *DDM1*-encoded protein. The full length protein or fragments of the protein may be used to advantage to generate an array of monoclonal or polyclonal antibodies specific for various epitopes of the protein, thereby providing even greater sensitivity for detection of the protein in cells or tissue.

DDM1 gene products also may be useful as pharmaceutical agents if it is determined that *DDM1* loss of function plays a role in carcinogenesis, as mentioned above. The gene products could be administered as replacement therapy for persons having neoplasias associated with *DDM1* loss of function.

Polyclonal or monoclonal antibodies immunologically specific for *DDM1*-encoded proteins may be used in a variety of assays designed to detect and quantitate the protein. Such assays include, but are not limited to: (1) flow cytometric analysis; (2) immunochemical localization in cultured cells or tissues; and (3) immunoblot analysis (e.g., dot blot, Western blot) of extracts from various cells and tissues.

Polyclonal or monoclonal antibodies that immunospecifically interact with the polypeptide encoded by *DDM1* can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules.

The following specific examples are provided to illustrate embodiments of the invention. They are not intended to limit the scope of the invention in any way.

EXAMPLE 1**Map-Based Isolation of the
Arabidopsis thaliana DDM1 Gene****Construction of recombination breakpoint lines.**

The recombination breakpoint lines were assembled in the
F3 generation from a parental cross between YI *DDM1*
ABA/YI *ddm1-2* ABA (Columbia strain (Col)) and
yi DDM1 aba/yi DDM1 aba (Landsberg erecta strain
(La er)). The recessive *yi* mutation leads to a yellow
inflorescence. The recessive *aba* mutation causes a defect
in abscisic acid biosynthesis and a wilting phenotype.
Information on genetic markers and the *A. thaliana*
genetic map can be found at: [http://genome-
www.stanford.edu/Arabidopsis/](http://genome-www.stanford.edu/Arabidopsis/). Selfed seeds from F1
YI ddm1-2 ABA/*yi DDM1 aba* plants were collected and 135
F2 recombinants (*yi* ABA, yellow inflorescence, non-
wilting; or YI *aba*: green inflorescence, wilting) were
identified. Selfed seeds from 111 of the 135 recombinant
F2 individuals were planted to generate F3 tissue for
genomic DNA preparation. The genotype at the *DDM1* locus
was scored in the F3 generation by Southern blot analysis
using methylation-sensitive endonucleases as described
previously (Vongs, A., Kakutani, T., Martienssen, R.A. &
Richards, E.J. , *Science* 260: 1926-1928, 1993).

Molecular markers. Two of the molecular
markers shown in Figure 1 were available from the
Arabidopsis research community: g4510 (*Arabidopsis*
Biological Resource Center (ABRC) stock# CD2-38) and
mi335 (ABRC stock# CD3-288). The remainder of the
molecular markers shown in Figure 1 were developed in
accordance with the present invention. *st10D21Bam* is an

- 32 -

insert end subclone of the BAC (bacterial artificial chromosome) clone T10D21 constructed by complete cleavage with *Bam*HI and recircularization. sT10D21Bam recognizes a Col/La er *Pst*I RFLP (restriction fragment length polymorphism). Molecular marker A is an *Xba*I Col/La er RFLP marker recognized by a 5.7 kb *Hind*III fragment of the C38 cosmid insert. Marker B is a *Rsa*I Col/La er CAPS marker (Koneieczny & Ausubel, Plant J. 4: 403-410, 1993) (forward primer: 5'-TCAAGGAGATGATTCGGGCGT-3', SEQ ID NO: 6; reverse primer: 5'-AAAGGACCCATTTACAGAACAC-3', SEQ ID NO:7). The remaining markers, C and D, correspond to RFLP's (*Bcl*I and *Pst*I, respectively) recognized by the succinate dehydrogenase cDNA clone, 105N23T7 (ABRC stock# 105N23T7).

Genomic library construction and screening. We screened the available *A. thaliana* BAC genomic libraries by standard colony hybridization techniques using radiolabeled 105N23T7 insert as a probe. The clone we subsequently focused upon, T10D21, came from the Texas A&M University BAC library (Choi et al., *Weeds World* 2: 17-20, 1995). To facilitate subsequent analysis, we cloned *Sau*3AI partially digested fragments from the T10D21 insert into the *Bam*HI site of SuperCos (Stratagene). We chose to further characterize one member of the resulting cosmid sublibrary, C38, which contained genetic markers that flanked *ddm1-2*. The C38 cosmid was submitted on April 20, 1999, under the provisions of the Budapest Treaty, with the American Type Culture Collection (Manassas VA), and assigned ATCC Accession No. 207208.

- 33 -

EXAMPLE 2

**DDM1 Gene Structure and Identification;
Sequence Determination of DDM1 Gene**

5 **DNA sequence determination.** C38 cosmid (~45 kb)
DNA, prepared using Qiagen columns and protocols, was
sonicated and 1-2 kb fragments isolated from a low-
melting temperature agarose gel. The size-selected DNA
was cloned into the *Sma*I site of a M13mp18 vector to
10 generate a shotgun library suitable for DNA sequence
determination. Single-stranded substrates were prepared
and sequenced using conventional dye-terminator cycle
sequencing protocols (Perkin-Elmer) on either an ABI 373
or ABI 377 automated DNA sequencer. The DNA sequence of
15 the *ddm1* alleles was determined using PCR-amplified
templates and oligonucleotide primers dispersed
throughout the *DDM1* gene. Sequence assembly and analysis
were accomplished using Phred/Phrap/Consed
(<http://www.mbt.washington.edu/>) and DNASTAR software
20 suites.

RT-PCR cDNA analysis. *DDM1* gene structure was
determined by analysis of the genomic DNA sequence and
the nucleotide sequence of RT-PCR (reverse transcription-
polymerase chain reaction) products encompassing the
25 coding region. *DDM1* and *ddm1-2* transcripts were analyzed
by RT-PCR as follows. Total RNA was prepared using the
Qiagen RNeasy™ protocol. Poly(A)+ transcripts were
collected on oligo-d(T)₂₅ magnetic Dynabeads (Dynal) and
first-strand cDNA synthesis performed following Dynal
30 protocols using Stratascript (Stratagene) reverse
transcriptase. Aliquots of the bead-immobilized first-
strand cDNA library were used as templates for PCR
amplification using KlenTaqI polymerase (Clontech). The
following oligonucleotide primers were used for the RT-
35 PCR experiment shown in Fig. 2b: forward,
5'-GCTGGAAGGGAAAGCTTAACAACC-3' (SEQ ID NO:8); reverse,

- 34 -

5'-ACACTGCCATCGATTCTGCAAACC-3' (SEQ ID NO:9).

GenBank accession numbers and SEQ ID NOS.

Arabidopsis DDM1 genomic DNA sequence: SEQ ID NO:1;

Arabidopsis DDM1 deduced amino acid sequence: SEQ ID NO:2;

5 *Arabidopsis* DDM1 5' upstream genomic DNA sequence: SEQ ID NO:3;

Mus musculus lymphocyte specific helicase (LSH); Genbank Accession No. AAB08015; SEQ ID NO:4;

10 *Homo sapiens* SNF2h; Genbank Accession No. AB010882; SEQ ID NO:5;

succinate dehydrogenase cDNA 105N23T7, T22529;

primers: SEQ ID NOS: 6-9.

15 While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set
20 forth in the following claims.

- 35 -

SEQUENCE LISTING

<110> Eric J. Richards
Jeffrey A. Jeddloh

<120> Plant Gene that Regulates DNA
Methylation

<130> WashU CI-0014PCT

<150> US 60/_____
<151> 1998-04-30

<150> US 09/104,070
<151> 1998-06-24

<160> 9

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 5000
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (535)...(566)

<221> CDS
<222> (772)...(850)

<221> CDS
<222> (986)...(1252)

<221> CDS
<222> (1354)...(1440)

<221> CDS
<222> (1549)...(1895)

<221> CDS
<222> (1976)...(2165)

<221> CDS
<222> (2251)...(2426)

<221> CDS
<222> (2559)...(2625)

<221> CDS
<222> (2703)...(2892)

<221> CDS
<222> (2975)...(3070)

<221> CDS
<222> (3148)...(3242)

<221> CDS
<222> (3317)...(3436)

- 36 -

```

<221> CDS
<222> (3540)...(3659)

<221> CDS
<222> (3745)...(3843)

<221> CDS
<222> (3934)...(4038)

<221> CDS
<222> (4130)...(4354)

<221> gene
<222> (535)...(4354)
<223> /gene="DDM1"

<221> mutation
<222> (785)...(785)
<223> /note= "site of ddml-5 (som8) mutation; delete G at 785 and
replace with 82 bp"

<221> mutation
<222> (2384)...(2385)
<223> /note= "site of ddml-6 (som4) mutation; delete G at 2384 or
2385"

<221> misc_feature
<222> (3186)...(3186)
<223> /note= "alternate splice donor site used in ddml-2"

<221> mutation
<222> (3243)...(3243)
<223> /note= "site of ddml-2 mutation; G to A"

<221> mutation
<222> (3337)...(3337)
<223> /note= "site of ddml-7 (som5) mutation; G to A"

<221> tRNA
<222> (4755)...(4826)
<223> /note= "complement of predicted tRNA-glu"

<400> 1
tgatcatttt cttcctccgg ccaatttgca gatcgaaaaa tgatttagct ttttattaaa      60
aatattgtta ttcgttttta gccgatatca taactttttg agatacatta tcaacacact      120
cgtgcaactg agatattctt gacacaattt ttgcatttga aattggcaat tttgtactac      180
tcatatagtt tgaagcttca attcactaca aaggttatta ctaattgtgt cgacaaatcc      240
agcagattta ataatgcccc ttccattaaa tgtttttttag ttttaataata ggatgatcat      300
atgacccaaa tcgtaaataa ggggttagggg taaacctgtc atttcaagct tcccgcccat      360
gggcgctact cccaatttaa taaaaaataa gaaaataggc gtaaatatga gagtgtgttt      420
tttcaatata cctcgggttt tgaatttgct ctcaaaagcg acggagacga ctgtttggct      480
cggtgatttc tcccgcggtt tgggtttttc ttaccggaat ttccttctcc ttcgatgggt      540
agtctgcgct ccagaaaagt tattccgtaa gtccctccac ctttccctttt catttcgtta      600
tttccggcga ttttctaggt ccttaacgct ctcgaaatcg ctcgctgttc ttggtgggtt      660
ttggttccct ctctgcgtaa ttttgtttgt cgtgtttttg gattatattc tctgactatt      720
ggtctcactg ttgatttatc atttctcgat tttggatttt tggactctta gggcttcgga      780
aatggtcagc gacgggaaaa cggagaaaga tgcgtctggt gattcacca cttctgttct      840
caacgaagag gtttgttcta tgttctacta ttttgccctc gtagtggtgt tgctttgtga      900
aactttgtgt gttactcttt gtttctttta atctggggtg ttctgtaa at gggtcctttt      960
tggtcctttt tttctgaatg tgaaggaaaa ctgtgaggag aaaagtgtta ctggtgtaga      1020
ggaagagata cttctagcca aaaatggaga ttcttctctt atttctgaag ccatggctca      1080
ggaggaagag cagctgctca aacttcggga agatgaagag aaagctaaca atgctggatc      1140

```

tgctgttgct	cctaattctga	atgaaactca	gtttactaaa	cttgatgagc	tcttgacgca	1200
aactcagctc	tactctgagt	ttctccttga	gaaaatggag	gatatcacaa	ttgtaattctt	1260
ctttattttct	ttcttctttg	tgggtttctca	cttttcgaat	gggagtcatt	attcttagtt	1320
tgaacaactt	gtgggtgaaa	tttgttttgc	tagaatggga	tagaaagtga	gagccaaaaa	1380
gctgagcccg	agaagactgg	tcgtggacgc	aaaagaaagg	ctgcttctca	gtacaacaat	1440
gttgggtcca	tttatataat	tttcaactac	tatgcatgat	cttgatatata	ttgttttttc	1500
tgcttggttg	agaaagtaac	ttacttggat	gcttttttct	tcaatcagac	taaggctaag	1560
agagcggttg	ctgctatgat	ttcaagatct	aaagaagatg	gtgagaccat	caactcagat	1620
ctgacagagg	aagaaacagt	catcaaaactg	cagaatgaac	tttgctctct	tctcactggg	1680
ggacagttaa	agtcttatca	gcttaaagg	gtcaaatggc	taatatcatt	gtggcagaat	1740
ggtttgaatg	gaatattagc	tgatcaaatg	ggacttggaa	agacgattca	aacgatcggt	1800
ttcttatcac	atctgaaagg	gaatgggttg	gatgggtccat	atctagtcac	tgctccactg	1860
tctacacttt	caaattgggt	caatgagatt	gctaggtact	ctcatggcca	tatgtgtttg	1920
tatagatcca	atgctttggg	gtttctgttg	aaagttttct	taccttttcc	attaggttca	1980
cgcttccat	caatgcaatc	atctaccatg	gggataaaaa	tcaaagggat	gagctcagga	2040
ggaagcacat	gcctaaaact	gttgggtccca	agttccctat	agttattact	tcttatgagg	2100
ttgccatgaa	tgatgctaaa	agaattctgc	ggcactatcc	atggaaatat	gttggtgattg	2160
atgaggtaaa	ttccgagatt	ggccaatgta	ctaggctttg	aagatcaaga	tgatctctct	2220
aactgataat	tttgttcttg	tatattatag	ggccacaggt	tgaaaaacca	caagtgtaaa	2280
ttgttgaggg	aactaaaaca	cttgaagatg	gataacaaac	ttctgctgac	aggaacacct	2340
ctgcaaaata	atctttctga	gctttgggtc	ttgttaaatt	ttattctgac	tgacatcttt	2400
acatcacatg	atgaatttga	atcatgggtac	aaacatgggtc	cttttctact	attatcccta	2460
actagtcttc	tttttttttt	tttttttggt	aacactgggtg	gcagcttttt	gacattttatt	2520
cctttcttag	tatctaactg	atagatgagt	ctctacaggt	ttgatttttc	tgaaaagaac	2580
aaaaacgaa	caaccaagga	agaagaagag	aaaagaagag	ctcaagtatg	tacaattata	2640
tcaattttcc	tttatttctt	tgattgtatt	tatgtcttat	gctaagggtg	catcttgtct	2700
aggttggttc	caaacttcat	ggtatactac	gaccattcat	ccttcgaaga	atgaaattgtg	2760
atgttgagct	ctcacttcca	cggaaaaagg	agattataat	gtatgctaca	atgactgata	2820
atcagaaaaa	gttccaggaa	catctgggtg	ataacacgtt	ggaagcacat	cttgagagaga	2880
atgccatccg	aggtacatga	tctatttttt	ttttttaata	ctttgtttta	ttatgtcatt	2940
ttctgcattc	atgtgttcac	cccctatact	tccaggtcaag	gctggaagg	aaagcttaac	3000
aacctgggtc	ttcaacttcg	aaagaactgc	aacctctcag	accttctcca	ggggcaataa	3060
gatgggtcat	gtatgtcagt	ttcttttaag	aaacgtaaga	aaaacttctg	tcatactgtt	3120
ctgtctaat	gtttcatttc	gtgacagatc	tctaccctcc	tggtgaagag	attgttggac	3180
agtgtggtaa	attccgctta	ttggagagat	tacttgttccg	gttattttgc	aataatcaca	3240
aagtatgttt	cacaaaacca	tggtcgttag	ctcatttccc	tttgagaact	tctctgatcc	3300
atttgcgtgat	gaccaggtcc	ttatcttctc	ccaattggacg	aaacttttgg	acattatgga	3360
ttactacttc	agtgagaagg	ggtttgaggt	ttgcagaatc	gatggcagtg	tgaagctgga	3420
tgaaggagga	agacaggttt	cacctgtgct	tatgctgctt	ttgcgttgct	tttaagcaat	3480
attctgacca	aatattataa	ccataagggtc	tctctctctc	tctctttgcc	ttgaaacaga	3540
ttaaagattt	cagtgatgag	aagagcagct	gtagtatatt	tctcctgagt	accagagctg	3600
gaggactcgg	aatcaatctt	actgctgctg	atacatgcat	cctctatgac	agcgtacagt	3660
taatcaaatc	aattaattta	ttttctttga	aggaaaatct	ttctctttcg	tggtgtctcc	3720
aactgtgttt	tgtctgatct	ccagaaccct	caaattggact	tgcaagccat	ggacagatgc	3780
cacagaatcg	ggcagacgaa	acctgttcat	gtttataggg	tttccacggc	tcagtcgata	3840
gaggtaaaac	tctttgttgt	tcatatcaat	caatcttaac	ttcaaaccat	tgagatttgtt	3900
gcctcatgag	attgggtttat	gacatttgct	cagaccggg	ttctgaaacg	agcgtacagt	3960
aagctcaagc	tggaacatgt	ggttattggc	caagggcaggt	ttcatcaaga	acgtgccaa	4020
tcttcaacac	ctttagaggt	tttaacttct	cttaaagctc	aatccttttt	agatacactt	4080
attatcaaca	aaatctccta	ttgacagctt	gaaccaaaact	aacacacagg	aagaggacat	4140
actggcggtg	cttaagggaag	atgaaactgc	tgaagataag	ttgatacaaa	ccgatataag	4200
cgatgcggat	cttgacaggt	tacttgaccg	gagtgcctg	acaatttactg	accggggaga	4260
gacacaagct	gctgaagctt	ttccagtga	gggtccaggt	tggaagtggt	tcctgcctag	4320
ttcggggagga	atgctgtctt	ccctgaacag	ttaggacaca	tttaataagcc	aggccttgaa	4380
accacttctg	tgtttttttt	ttttttttcc	ggaacatgat	cggttacttt	tggtctgggag	4440
gatttaatta	ttagaggggt	cggaagtgtt	tgtaagttaa	agaactcact	taaaaccctg	4500
aaaacatgac	agttaatggg	gattagctct	caatgtgatg	aaaacaattg	gcccctctgat	4560
tttgctgttg	cggtaatatt	atgacttgtg	tacgtttata	gtctttgtag	tctgcaatttt	4620
tggtcattgag	ctattttctca	cgaacttatg	ggatcttatg	ttttggatttt	gggattttgtt	4680
aacttatatg	attaggctca	atagtttcac	agaatattaa	aaacttgagt	aggggtttaaa	4740
aaagaagcaa	aaagctccga	tgccgggaat	cgaaccggg	tctcctgggt	gaaagccaga	4800
tatcctaacc	gctggacgac	atcggatttg	ttgatgtcta	ttcttgtaaa	tagtaaatat	4860

- 38 -

ttagttttat cggtttttgca tctaattggac taaaacatga acacgagacg ccgacaagaa 4920
 tgaatggggc aggcacccaaa catttgggta aaagtatgca gtgggggtatt attgacaatt 4980
 tgaccattac aagagctaatt 5000

<210> 2
 <211> 764
 <212> PRT
 <213> Arabidopsis thaliana

<400> 2
 Met Val Ser Leu Arg Ser Arg Lys Val Ile Pro Ala Ser Glu Met Val
 1 5 10 15
 Ser Asp Gly Lys Thr Glu Lys Asp Ala Ser Gly Asp Ser Pro Thr Ser
 20 25 30
 Val Leu Asn Glu Glu Glu Asn Cys Glu Glu Lys Ser Val Thr Val Val
 35 40 45
 Glu Glu Glu Ile Leu Leu Ala Lys Asn Gly Asp Ser Ser Leu Ile Ser
 50 55 60
 Glu Ala Met Ala Gln Glu Glu Gln Leu Leu Lys Leu Arg Glu Asp
 65 70 75 80
 Glu Glu Lys Ala Asn Asn Ala Gly Ser Ala Val Ala Pro Asn Leu Asn
 85 90 95
 Glu Thr Gln Phe Thr Lys Leu Asp Glu Leu Leu Thr Gln Thr Gln Leu
 100 105 110
 Tyr Ser Glu Phe Leu Leu Glu Lys Met Glu Asp Ile Thr Ile Asn Gly
 115 120 125
 Ile Glu Ser Glu Ser Gln Lys Ala Glu Pro Glu Lys Thr Gly Arg Gly
 130 135 140
 Arg Lys Arg Lys Ala Ala Ser Gln Tyr Asn Asn Thr Lys Ala Lys Arg
 145 150 155 160
 Ala Val Ala Ala Met Ile Ser Arg Ser Lys Glu Asp Gly Glu Thr Ile
 165 170 175
 Asn Ser Asp Leu Thr Glu Glu Glu Thr Val Ile Lys Leu Gln Asn Glu
 180 185 190
 Leu Cys Pro Leu Leu Thr Gly Gly Gln Leu Lys Ser Tyr Gln Leu Lys
 195 200 205
 Gly Val Lys Trp Leu Ile Ser Leu Trp Gln Asn Gly Leu Asn Gly Ile
 210 215 220
 Leu Ala Asp Gln Met Gly Leu Gly Lys Thr Ile Gln Thr Ile Gly Phe
 225 230 235 240
 Leu Ser His Leu Lys Gly Asn Gly Leu Asp Gly Pro Tyr Leu Val Ile
 245 250 255
 Ala Pro Leu Ser Thr Leu Ser Asn Trp Phe Asn Glu Ile Ala Arg Phe
 260 265 270
 Thr Pro Ser Ile Asn Ala Ile Ile Tyr His Gly Asp Lys Asn Gln Arg
 275 280 285
 Asp Glu Leu Arg Arg Lys His Met Pro Lys Thr Val Gly Pro Lys Phe
 290 295 300
 Pro Ile Val Ile Thr Ser Tyr Glu Val Ala Met Asn Asp Ala Lys Arg
 305 310 315 320
 Ile Leu Arg His Tyr Pro Trp Lys Tyr Val Val Ile Asp Glu Gly His
 325 330 335
 Arg Leu Lys Asn His Lys Cys Lys Leu Arg Glu Leu Lys His Leu
 340 345 350
 Lys Met Asp Asn Lys Leu Leu Leu Thr Gly Thr Pro Leu Gln Asn Asn
 355 360 365
 Leu Ser Glu Leu Trp Ser Leu Leu Asn Phe Ile Leu Pro Asp Ile Phe
 370 375 380
 Thr Ser His Asp Glu Phe Glu Ser Trp Phe Asp Phe Ser Glu Lys Asn
 385 390 395 400
 Lys Asn Glu Ala Thr Lys Glu Glu Glu Glu Lys Arg Arg Ala Gln Val
 405 410 415

- 39 -

Val Ser Lys Leu His Gly Ile Leu Arg Pro Phe Ile Leu Arg Arg Met
 420 425 430
 Lys Cys Asp Val Glu Leu Ser Leu Pro Arg Lys Lys Glu Ile Ile Met
 435 440 445
 Tyr Ala Thr Met Thr Asp His Gln Lys Lys Phe Gln Glu His Leu Val
 450 455 460
 Asn Asn Thr Leu Glu Ala His Leu Gly Glu Asn Ala Ile Arg Gly Gln
 465 470 475 480
 Gly Trp Lys Gly Lys Leu Asn Asn Leu Val Ile Gln Leu Arg Lys Asn
 485 490 495
 Cys Asn His Pro Asp Leu Leu Gln Gly Gln Ile Asp Gly Ser Tyr Leu
 500 505 510
 Tyr Pro Pro Val Glu Glu Ile Val Gly Gln Cys Gly Lys Phe Arg Leu
 515 520 525
 Leu Glu Arg Leu Leu Val Arg Leu Phe Ala Asn Asn His Lys Val Leu
 530 535 540
 Ile Phe Ser Gln Trp Thr Lys Leu Leu Asp Ile Met Asp Tyr Tyr Phe
 545 550 555 560
 Ser Glu Lys Gly Phe Glu Val Cys Arg Ile Asp Gly Ser Val Lys Leu
 565 570 575
 Asp Glu Arg Arg Arg Gln Ile Lys Asp Phe Ser Asp Glu Lys Ser Ser
 580 585 590
 Cys Ser Ile Phe Leu Leu Ser Thr Arg Ala Gly Gly Leu Gly Ile Asn
 595 600 605
 Leu Thr Ala Ala Asp Thr Cys Ile Leu Tyr Asp Ser Asp Trp Asn Pro
 610 615 620
 Gln Met Asp Leu Gln Ala Met Asp Arg Cys His Arg Ile Gly Gln Thr
 625 630 635 640
 Lys Pro Val His Val Tyr Arg Leu Ser Thr Ala Gln Ser Ile Glu Thr
 645 650 655
 Arg Val Leu Lys Arg Ala Tyr Ser Lys Leu Lys Leu Glu His Val Val
 660 665 670
 Ile Gly Gln Gly Gln Phe His Gln Glu Arg Ala Lys Ser Ser Thr Pro
 675 680 685
 Leu Glu Glu Glu Asp Ile Leu Ala Leu Leu Lys Glu Asp Glu Thr Ala
 690 695 700
 Glu Asp Lys Leu Ile Gln Thr Asp Ile Ser Asp Ala Asp Leu Asp Arg
 705 710 715 720
 Leu Leu Asp Arg Ser Asp Leu Thr Ile Thr Ala Pro Gly Glu Thr Gln
 725 730 735
 Ala Ala Glu Ala Phe Pro Val Lys Gly Pro Gly Trp Glu Val Val Leu
 740 745 750
 Pro Ser Ser Gly Gly Met Leu Ser Ser Leu Asn Ser
 755 760

<210> 3

<211> 5000

<212> DNA

<213> Arabidopsis thaliana

<400> 3

tgctgaagtt	tccatggaag	attgtgacca	cgacgatgaa	gctgaagatt	ctgggtcacgt	60
tgaaaacctt	tgttacagat	ttcgcaaacg	aatcgattcg	ttgccataag	tgtttttaggt	120
gacaaaagcta	tcacttcagc	gtctggatct	gaatttagac	aatcagtgag	aacaactaaa	180
aacagaaaat	ttcaaaactca	aaaaacagaa	aaaaaaaaagt	ttggattttt	gagaagtacc	240
aggcattcca	ggaagattcc	gtttcttctt	cccgacggat	ttaggagtta	gatttttggtt	300
tccggtcgat	gagacgcttg	catcgccgga	aactgtagag	gaattatcta	aatcaaccgg	360
catgtttcaa	agataactaaa	ttccaatctt	tgaacacaaa	aaggaagaag	caaattctcag	420
ctcagctcaa	tctaggggtt	atcatcctcc	tctactctg	tttagtctct	ctttctctct	480
ctctttcttca	gctaccagtc	aatctgcttt	tcgtaaaaat	ctccttttcc	cctttccgcc	540
accaaaacttt	tctgataact	cactctctga	cctctcttct	tcaaaaagat	ttaaaacccc	600
caaaagaaaa	agaaaaaaa	tcaaaaacttc	attacccaag	aaatctctta	atcatttaac	660

ccagactctt	tcttctccac	acgcactctt	tatccaccgt	ccaccgatct	gatccaacgg	720
ctgagatttc	accggagacg	agttatcctt	actacttccg	gcttgtttct	ctctgaagaa	780
tcaccggaaa	aaaaataaag	gcggcttggt	tgtgagactt	tgtgtgaaag	cttcaacctt	840
ttttttcttt	ttctttggct	tgtccaagaa	aaaggagcct	tcttcttctt	ttctctctct	900
ggagacaatt	atactaattt	ttttcttttc	aactttttcac	cctttttttt	ttgttaacaa	960
acatttttta	tacataattg	tgtcgacttt	caagttccaa	gtatctaaat	ctgtattttg	1020
gactcccatg	caaataatta	aaatagaata	atctttttgt	agatttttaa	ttgaaaacgg	1080
tgtagaaagg	ttaaaagcac	caaacaaaac	gagtaaatag	atattgtaat	aattttttca	1140
ccttttatgga	aaagattata	tcatagacga	tgtacacaga	tgaaaattag	aaaatggcat	1200
gtgaatatat	gcagtaccca	atgaatgcaa	tatcagggtt	gtattatttt	tctattgtat	1260
ctctacatgt	tacgtaatca	aacgatcaag	taattttatta	atattgtcga	tggcgtagaa	1320
attataaatt	tattttatgt	cattgtttac	tatatagatt	ttgagctaaa	cgacttattt	1380
tgtcaaaaga	tatatccgtg	tttggtttaa	gattgggttt	tagtattttc	aatattaatc	1440
taaattctta	gcttatgaac	atgtcaataa	acaaaaaaat	tattttactg	tcactgtcct	1500
tagacgggga	caaaggaggg	tattaccgtc	gcgttgtcgg	accgtaaaaa	aattaaacca	1560
attttgttgt	tgaacgaata	acatttttta	ctgtgggaat	ttgtcgtgta	gcattacgtt	1620
cgaaatcgca	atttgttttc	ttctttgtgg	gtgtatatatt	ctgggttaacg	aaactataac	1680
ccaatttaat	gcaatgttcg	tctgtttttg	ttgactttga	cccttttttg	gtaatatctg	1740
ttcagctttt	gttttaacgt	tttcattgcc	ttgtaggcat	ctgagaagct	cagattctga	1800
cacgtgtctt	ttgttatctg	aatttgcata	cgttggataa	acatgacgct	gacagggtga	1860
ttgaaaagta	accagcttgg	atctctgtgt	atatgttaca	ccgccacttc	ccttaatttc	1920
ttcgttctta	gttaaaataa	aaaagggtta	atttatgagt	aaaagtatgt	aaaacgacaa	1980
cgattactat	aagaatttaa	atttatcttt	gcttagtaat	ttgcacttaa	gattggattc	2040
aaattttgta	aaaagcgaat	gttacatata	tgtccattga	aaaaattgca	tttgacttta	2100
caagcattga	aattaattaa	tttgggaccc	ctttttttgt	tagtttcaaa	ggaagaatta	2160
ttttaggctg	agatgggtcc	ctccataaac	tcactattct	gccagcatac	aaattcctta	2220
acatatgggt	caaatagcag	ttccaaccac	tagtatccaa	taataatctg	aacaaattat	2280
ctttcttttt	tttccctgat	aatcttgat	ttgtttgttc	aatgagctta	atacgtatat	2340
tagttatgac	ttataactaa	atactttgac	tcacttgatc	cgtaacacatt	gatttcgttt	2400
attcaaatcc	gaacaacgta	atgatctttt	tggggccgagt	tatttgtatt	ctcaacctga	2460
gtccaaccat	gctttatggg	cttttctgtt	tatttatgca	tgtaaagtgt	ataatgcttg	2520
caaataacca	catattgtat	gaatgtaatt	actatgattt	aagggcactg	cttttctgtt	2580
ttcacgttgt	tttcgaaatt	gctattgcgt	gtgatattct	tgttggacca	attattgaaa	2640
aggacaaggc	tgactctggg	ttttaatgag	tagtcccat	gggagttatg	ttcatttacc	2700
acacattttt	ttgtatagta	tagtatgagt	ttttatttga	tatcttttat	cttcggaaaa	2760
taaatgggtc	aaattgtttg	tctaaaaatg	cacacatgaa	tatcttgttg	tctcacacaa	2820
ttgtaggaaa	caaattaata	tttgttgcca	aaataatgtt	attattttat	catacgaat	2880
cctagagaaa	atgggtggcaa	aagaggcaaa	gactaaacta	atgaatttaa	aatatgaaaa	2940
tgatggaatg	actgggtttac	caatattaca	gtatatgtga	attttataaa	aacgaatcct	3000
gaagaagagg	gcaaacccca	agaccacgca	aatcagtcta	caaatatgaa	aattttccat	3060
aactagaaaa	acatgtgcat	ttatcttttt	ccatcattcg	gattttttaca	atggaaattt	3120
tgaccactga	gcgcaagtgt	tatagtattt	tattattatc	caatattaat	atcattattc	3180
ggatccatgc	attctatata	actatgtcca	ccatcttact	tgtgtctatg	ttgcaacttc	3240
aacgtcgtat	atataatagg	attgttgtca	cgaatacaat	gctaattaa	gaagatttgt	3300
acttctcgga	aaatttagaa	ctaattaaga	gtggaactaa	aatgccaatg	aaaaatgcct	3360
aaatcaaagg	agaaccacaa	atataaattg	gaagacctta	aaaaacaatt	aaacgaggac	3420
gaaacaaatt	ttggaatcat	caattatacg	aaaaaaaagaa	gaaagaaaaa	agagggtttca	3480
tgaatcacag	tagtgctgac	aatcttcgaa	ccatttgttg	gtttcatata	atcgatcacc	3540
aatagaacaa	aagagaaaaca	gaggaacaga	aagaatagaa	ggagtgggaa	gtgtatgagg	3600
aagctgtgtc	cgaacataga	caaagacgat	ggctcggaga	cggtgttgga	agttccgata	3660
ccggaggaga	tgttttccgg	tatgggcaac	aacgttgcac	ttagggtggca	aaatatgatg	3720
acgtggatga	aagctcaaac	gtctgataaa	tggctcgcaac	cgcttatcgc	cgctcgtatc	3780
aacgagctcc	ggttccttct	ctacctcggt	ggctcgccct	ttatacctct	ccagggttcaa	3840
gtcggtcact	ctgttcataa	gcccgtcaaa	gattgtcaca	ttgtaagtca	ttcaaaatca	3900
atccttatga	aaacataaca	aagatgttga	aaatagtgatt	cctctttttt	tttctttttt	3960
ttctttttatg	atcaaaaacc	aaaaaagtca	ttaccttgct	tcgtaagtat	tcaacataaa	4020
gttggttaate	catgtgttgt	actctgcaag	tctgcattac	attattcatc	gtacacagag	4080
tcatacaact	cagtttcatt	gtttttttgc	ttatgaatta	cgattgcagc	aagcttcaac	4140
ggcgaataac	attgtacagc	agtacatagc	agcgacggga	ggaccacagg	cgtaaacgcg	4200
cgtgaacagc	atgtgcgtca	cgggacaagt	gaagatgacg	gcgtcggagt	ttcatcaagg	4260
agatgattcg	ggcgttaate	taaagagcaa	cgacgaaatg	gggtgggttcg	ttttatggca	4320
aaaggatcca	gatcttttgg	gtttggagct	cgctcgtctc	gggtgcaaa	tggatatgtg	4380

- 41 -

```

gaagcaacgg tcggctttca tggcgacatt cctctaacca gcaaactccg gcgtctacsg 4440
gaacgccaar acctctccgc cggtttwtac aggtccaatc cggttattga ttttttttk 4500
gatgtaatgt ccggttctca aaatgttgaa ccggtgggtt atttattgtt tggagcaggg 4560
gttaratcct cgttcgacgg cgaatctgtt tcttgacgca aacgtgtatc ggagagaaga 4620
taatcaacgg tgaggattgc tttatcttga aactggagac gagtccggcg gttcgagaag 4680
ctcaaagcgg tccgaatttt gagataattc atcacacgat atgggggttat tttagtcaaa 4740
gatcgggact tttgattcag ttcgaagatt cgcggctttt gagaatgagg accaaggaa 4800
acgaagatgt cttctgggag actagtgtct agtcgggtgat ggatgattac cgatacggtg 4860
acaatgtgaa catcgctcac ggcgggaaaa catcggtcac ggttttccgg tacggtgaag 4920
cgtcggcgaa tcacgggaga cagatgacgg agaagtggag gatagaagaa gttgatttta 4980
atgtttgggg tctctccgtt 5000

```

<210> 4
 <211> 603
 <212> PRT
 <213> Mus musculus

```

<400> 4
Met Leu Trp Glu Asn Gly Ile Asn Gly Ile Leu Ala Asp Glu Met Gly
1 5 10 15
Leu Gly Lys Thr Val Gln Cys Ile Ala Thr Ile Ala Leu Met Ile Gln
20 25 30
Arg Gly Val Pro Gly Pro Phe Leu Val Cys Gly Pro Leu Ser Thr Leu
35 40 45
Pro Asn Trp Met Ala Glu Phe Lys Arg Phe Thr Pro Glu Ile Pro Thr
50 55 60
Leu Leu Tyr His Gly Thr Arg Glu Asp Arg Arg Lys Leu Val Lys Asn
65 70 75 80
Ile His Lys Arg Gln Gly Thr Leu Gln Ile His Pro Val Val Val Thr
85 90 95
Ser Phe Glu Ile Ala Met Arg Asp Gln Asn Ala Leu Gln His Cys Tyr
100 105 110
Trp Lys Tyr Leu Ile Val Asp Glu Gly His Arg Ile Lys Asn Met Lys
115 120 125
Cys Arg Leu Ile Arg Glu Leu Lys Arg Phe Asn Ala Asp Asn Lys Leu
130 135 140
Leu Leu Thr Gly Thr Pro Leu Gln Asn Asn Leu Ser Glu Leu Trp Ser
145 150 155 160
Leu Leu Asn Phe Leu Leu Pro Asp Val Phe Asp Asp Leu Lys Ser Phe
165 170 175
Glu Ser Trp Phe Asp Ile Thr Ser Leu Ser Glu Thr Ala Glu Asp Ile
180 185 190
Ile Ala Lys Glu Arg Glu Gln Asn Val Leu His Met Leu His Gln Ile
195 200 205
Leu Thr Pro Phe Leu Leu Arg Arg Leu Lys Ser Asp Val Ala Leu Glu
210 215 220
Val Pro Pro Lys Arg Glu Val Val Val Tyr Ala Pro Leu Cys Asn Lys
225 230 235 240
Gln Glu Ile Phe Tyr Thr Ala Ile Val Asn Arg Thr Ile Ala Asn Met
245 250 255
Phe Gly Ser Cys Glu Lys Glu Thr Val Glu Leu Ser Pro Thr Gly Arg
260 265 270
Pro Lys Arg Arg Ser Arg Lys Ser Ile Asn Tyr Ser Glu Leu Asp Gln
275 280 285
Phe Pro Ser Glu Leu Glu Lys Leu Ile Ser Gln Ile Gln Pro Glu Val
290 295 300
Asn Arg Glu Arg Thr Val Val Glu Gly Asn Ile Pro Ile Glu Ser Glu
305 310 315 320
Val Asn Leu Lys Leu Arg Asn Ile Met Met Leu Leu Arg Lys Cys Cys
325 330 335
Asn His Pro Tyr Met Ile Glu Tyr Pro Ile Asp Pro Val Thr Gln Glu
340 345 350

```

- 42 -

Phe Lys Ile Asp Glu Glu Leu Val Thr Asn Ser Gly Lys Phe Leu Ile
 355 360 365
 Leu Asp Arg Met Leu Pro Glu Leu Lys Lys Arg Gly His Lys Val Leu
 370 375 380
 Val Phe Ser Gln Met Thr Ser Met Leu Asp Ile Leu Met Asp Tyr Cys
 385 390 395 400
 His Leu Arg Asn Phe Ile Phe Ser Arg Leu Asp Gly Ser Met Ser Tyr
 405 410 415
 Ser Glu Arg Glu Lys Asn Ile Tyr Ser Phe Asn Thr Asp Pro Asp Val
 420 425 430
 Phe Leu Phe Leu Val Ser Thr Arg Ala Gly Gly Leu Gly Ile Asn Leu
 435 440 445
 Thr Ala Ala Asp Thr Val Ile Ile Tyr Asp Ser Asp Trp Asn Pro Gln
 450 455 460
 Ser Asp Leu Gln Ala Gln Asp Arg Cys His Arg Ile Gly Gln Thr Lys
 465 470 475 480
 Pro Val Val Val Tyr Arg Leu Val Thr Ala Asn Thr Ile Asp Gln Lys
 485 490 495
 Ile Val Glu Arg Ala Ala Ala Lys Arg Lys Leu Glu Lys Leu Ile Ile
 500 505 510
 His Lys Asn His Phe Lys Gly Gly Gln Ser Gly Leu Ser Gln Ser Lys
 515 520 525
 Asn Phe Leu Asp Ala Lys Glu Leu Met Glu Leu Leu Lys Ser Arg Asp
 530 535 540
 Tyr Glu Arg Glu Val Lys Gly Ser Arg Glu Lys Val Ile Ser Asp Glu
 545 550 555 560
 Asp Leu Glu Leu Leu Asp Arg Ser Asp Leu Ile Asp Gln Met Lys
 565 570 575
 Ala Ser Arg Pro Ile Lys Gly Lys Thr Gly Ile Phe Lys Ile Leu Glu
 580 585 590
 Asn Ser Glu Asp Ser Ser Ala Glu Cys Leu Phe
 595 600

<210> 5
 <211> 1052
 <212> PRT
 <213> Homo sapiens

<400> 5
 Met Ser Ser Ala Ala Glu Pro Pro Pro Pro Pro Pro Glu Ser Ala
 1 5 10 15
 Pro Ser Lys Pro Ala Ala Ser Ile Ala Ser Gly Gly Ser Asn Ser Ser
 20 25 30
 Asn Lys Gly Gly Pro Glu Gly Val Ala Ala Gln Ala Val Ala Ser Ala
 35 40 45
 Ala Ser Ala Gly Pro Ala Asp Ala Glu Met Glu Glu Ile Phe Asp Asp
 50 55 60
 Ala Ser Pro Gly Lys Gln Lys Glu Ile Gln Glu Pro Asp Pro Thr Tyr
 65 70 75 80
 Glu Glu Lys Met Gln Thr Asp Arg Ala Asn Arg Phe Glu Tyr Leu Leu
 85 90 95
 Lys Gln Thr Glu Leu Phe Ala His Phe Ile Gln Pro Ala Ala Gln Lys
 100 105 110
 Thr Pro Thr Ser Pro Leu Lys Met Lys Pro Gly Arg Pro Arg Ile Lys
 115 120 125
 Lys Asp Glu Lys Gln Asn Leu Leu Ser Val Gly Asp Tyr Arg His Arg
 130 135 140
 Arg Thr Glu Gln Glu Glu Asp Glu Glu Leu Leu Thr Glu Ser Ser Lys
 145 150 155 160
 Ala Thr Asn Val Cys Thr Arg Phe Glu Asp Ser Pro Ser Tyr Val Lys
 165 170 175
 Trp Gly Lys Leu Arg Asp Tyr Gln Val Arg Gly Leu Asn Trp Leu Ile

BNSDOCID: <WO__9955891A1_IB>

- 44 -

```

        675          680          685
Asn Glu Lys Leu Ser Lys Met Gly Glu Ser Ser Leu Arg Asn Phe Thr
690          695          700
Met Asp Thr Glu Ser Ser Val Tyr Asn Phe Glu Gly Glu Asp Tyr Arg
705          710          715          720
Glu Lys Gln Lys Ile Ala Phe Thr Glu Trp Ile Glu Pro Pro Lys Arg
725          730          735
Glu Arg Lys Ala Asn Tyr Ala Val Asp Ala Tyr Phe Arg Glu Ala Leu
740          745          750
Arg Val Ser Glu Pro Lys Ala Pro Lys Ala Pro Arg Pro Pro Lys Gln
755          760          765
Pro Asn Val Gln Asp Phe Gln Phe Phe Pro Pro Arg Leu Phe Glu Leu
770          775          780
Leu Glu Lys Glu Ile Leu Phe Tyr Arg Lys Thr Ile Gly Tyr Lys Val
785          790          795          800
Pro Arg Asn Pro Glu Leu Pro Asn Ala Ala Gln Ala Gln Lys Glu Glu
805          810          815
Gln Leu Lys Ile Asp Glu Ala Glu Ser Leu Asn Asp Glu Glu Leu Glu
820          825          830
Glu Lys Glu Lys Leu Leu Thr Gln Gly Phe Thr Asn Trp Asn Lys Arg
835          840          845
Asp Phe Asn Gln Phe Ile Lys Ala Asn Glu Lys Trp Gly Arg Asp Asp
850          855          860
Ile Glu Asn Ile Ala Arg Glu Val Glu Gly Lys Thr Pro Glu Glu Val
865          870          875          880
Ile Glu Tyr Ser Ala Val Phe Trp Glu Arg Cys Asn Glu Leu Gln Asp
885          890          895
Ile Glu Lys Ile Met Ala Gln Ile Glu Arg Gly Glu Ala Arg Ile Gln
900          905          910
Arg Arg Ile Ser Ile Lys Lys Ala Leu Asp Thr Lys Ile Gly Arg Tyr
915          920          925
Lys Ala Pro Phe His Gln Leu Arg Ile Ser Tyr Gly Thr Asn Lys Gly
930          935          940
Lys Asn Tyr Thr Glu Glu Asp Arg Phe Leu Ile Cys Met Leu His
945          950          955          960
Lys Leu Gly Phe Asp Lys Glu Asn Val Tyr Asp Glu Leu Arg Gln Cys
965          970          975
Ile Arg Asn Ser Pro Gln Phe Arg Phe Asp Trp Phe Leu Lys Ser Arg
980          985          990
Thr Ala Met Glu Leu Gln Arg Arg Cys Asn Thr Leu Ile Thr Leu Ile
995          1000          1005
Glu Arg Glu Asn Met Glu Leu Glu Glu Lys Glu Lys Ala Glu Lys Lys
1010          1015          1020
Lys Arg Gly Pro Lys Pro Ser Thr Gln Lys Arg Lys Met Asp Gly Ala
1025          1030          1035          1040
Pro Asp Gly Arg Gly Arg Lys Lys Lys Leu Lys Leu
1045          1050

```

```

<210> 6
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<223> /note= "synthetic construct"

```

```

<400> 6
tcaaggagat gattcgggcg t

```

21

```

<210> 7
<211> 22
<212> DNA
<213> Artificial Sequence

```

- 45 -

<220>
<223> /note= "synthetic construct"

<400> 7
aaaggaccca ttacagaac ac 22

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<223> /note= "synthetic construct"

<400> 8
gctggaagg aaagcttaac aacc 24

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
<223> /note= "synthetic construct"

<400> 9
acactgccat cgattctgca aacc 24

We claim:

1. An isolated nucleic acid molecule comprising a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, said gene occupying a segment of said chromosome 5, lower arm, flanked on the centromeric side within 20 kilobases by a gene encoding a zinc-finger protein and on the telomeric side within 1 kilobase by a gene encoding a glutamic acid tRNA, the disruption of said gene being associated with DNA hypomethylation.
2. The nucleic acid molecule of claim 1, wherein said gene is composed of exons that form an open reading frame having a sequence that encodes a polypeptide about 750-850 amino acids in length.
3. A cDNA molecule comprising the exons of the nucleic acid molecule of claim 2.
4. The nucleic acid molecule of claim 2, wherein said open reading frame encodes an amino acid sequence substantially the same as SEQ ID NO:2.
5. The nucleic acid molecule of claim 4, wherein said open reading frame encodes amino acid SEQ ID NO:2.
6. The nucleic acid molecule of claim 5, which comprises an open reading frame of SEQ ID NO:1.
7. A recombinant DNA molecule, comprising a vector having an insert that includes the nucleic acid molecule of claim 1.
8. The recombinant DNA molecule of claim 7,

- 47 -

which is cosmid C38, ATCC Accession No. 207208.

9. An oligonucleotide between about 10 and 100 nucleotides in length, which specifically hybridizes with
5 a portion of the nucleic acid molecule of claim 1.

10. An isolated nucleic acid molecule which is a gene, the disruption of which is associated with DNA hypomethylation, having a sequence selected from the
10 group consisting of:

a) SEQ ID NO:1;

b) an allelic variant or natural mutant of
SEQ ID NO:1;

c) a sequence hybridizing with part or
15 all of SEQ ID NO:1 or its complement and encoding a polypeptide substantially the same as part or all of a polypeptide encoded by SEQ ID NO:1;

d) a sequence encoding part or all of a polypeptide having amino acid SEQ ID NO:2; and

20 e) a sequence encoding part or all of a polypeptide contained in the cosmid clone C38, designated ATCC Accession No. 207208.

11. A polypeptide produced by expression of an
25 isolated nucleic acid molecule comprising part or all of an open reading frame of a gene located on *Arabidopsis thaliana* chromosome 5, lower arm, said gene occupying a segment of said chromosome 5, lower arm, flanked on the centromeric side within 20 kilobases by a gene encoding a
30 zinc-finger protein and on the telomeric side within 1 kilobase by a gene encoding a glutamic acid tRNA, the disruption of said gene being associated with DNA hypomethylation.

35 12. The polypeptide of claim 11, produced by

- 48 -

expression of a sequence selected from the group consisting of:

- a) SEQ ID NO:1;
- b) an allelic variant or natural mutant of
5 SEQ ID NO:1;
- c) a sequence hybridizing with part or all of SEQ ID NO:1 or its complement and encoding a polypeptide substantially the same as part or all of a polypeptide encoded by SEQ ID NO:1;
- 10 d) a sequence encoding part or all of a polypeptide having amino acid SEQ ID NO:2; and
- e) a sequence encoding part or all of a polypeptide contained in the clone designated ATCC Accession No. 207208.

15

13. The polypeptide of claim 11, having the amino acid sequence of part or all of SEQ ID NO:2.

14. An antibody immunologically specific for
20 the polypeptide of claim 11.

15. An isolated nucleic acid molecule having a sequence substantially the same as SEQ ID NO:3.

25 16. An isolated protein encoded by an *Arabidopsis thaliana* gene, said protein being a member of an SWI2/SNF2 family of polypeptides, loss of function of said protein being associated with DNA hypomethylation.

30 17. The protein of claim 16, encoded by a gene located on *A. thaliana* chromosome 5, lower arm, centromerically flanked within 20 kilobases by a zinc-finger-encoding gene and telomerically within one kilobase by a gene encoding a glutamic acid tRNA.

35

- 49 -

18. The protein of claim 16, encoded by a DNA segment on a recombinant cosmid C38, having ATCC Accession No. 207208.

5

19. The protein of claim 16, having amino acid SEQ ID NO:2.

20. A transgenic organism comprising the nucleic acid molecule of claim 1.

10

21. The transgenic organism of claim 20, which is a plant.

15

22. A method of stabilizing fidelity of DNA methylation in an organism, comprising transforming the organism with the nucleic acid molecule of claim 1.

23. A method of reducing or eliminating gene silencing in a plant, comprising inhibiting or preventing expression of an endogenous *DDM1* gene of the plant.

20

24. A method of introducing inbreeding depression in a plant, comprising inhibiting or preventing expression of an endogenous *DDM1* gene of the plant.

25

1/4

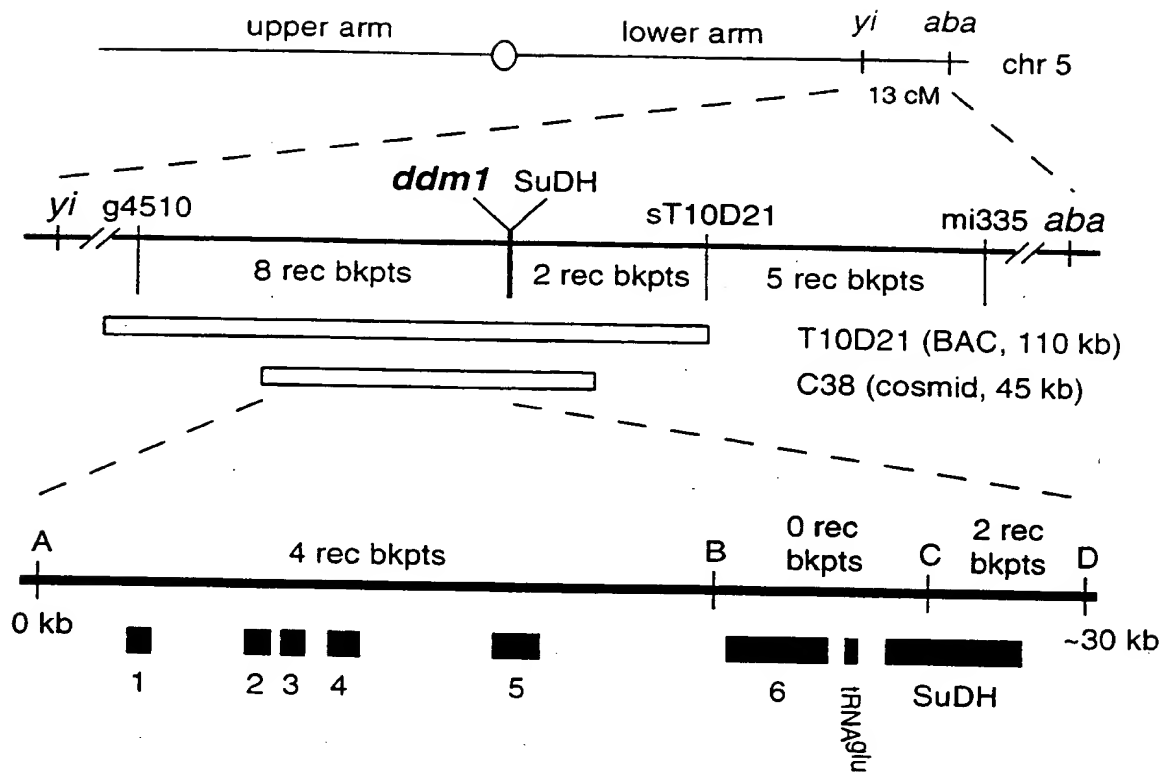


Figure 1

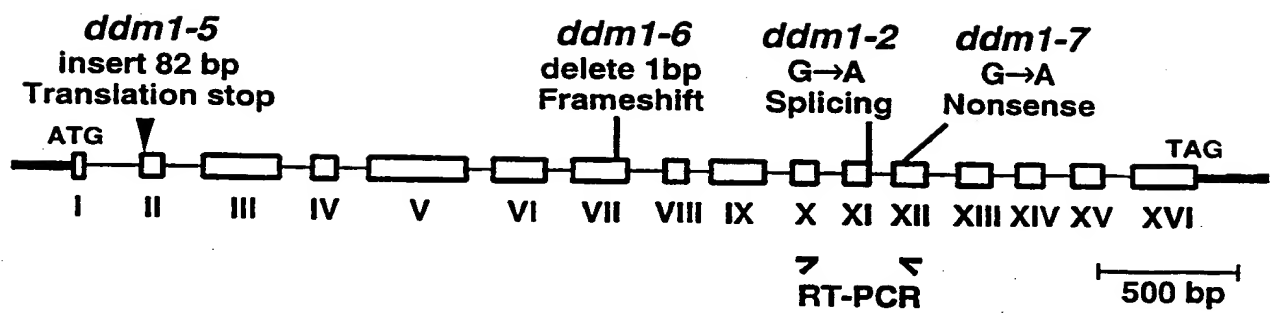


Fig. 2A

3/4

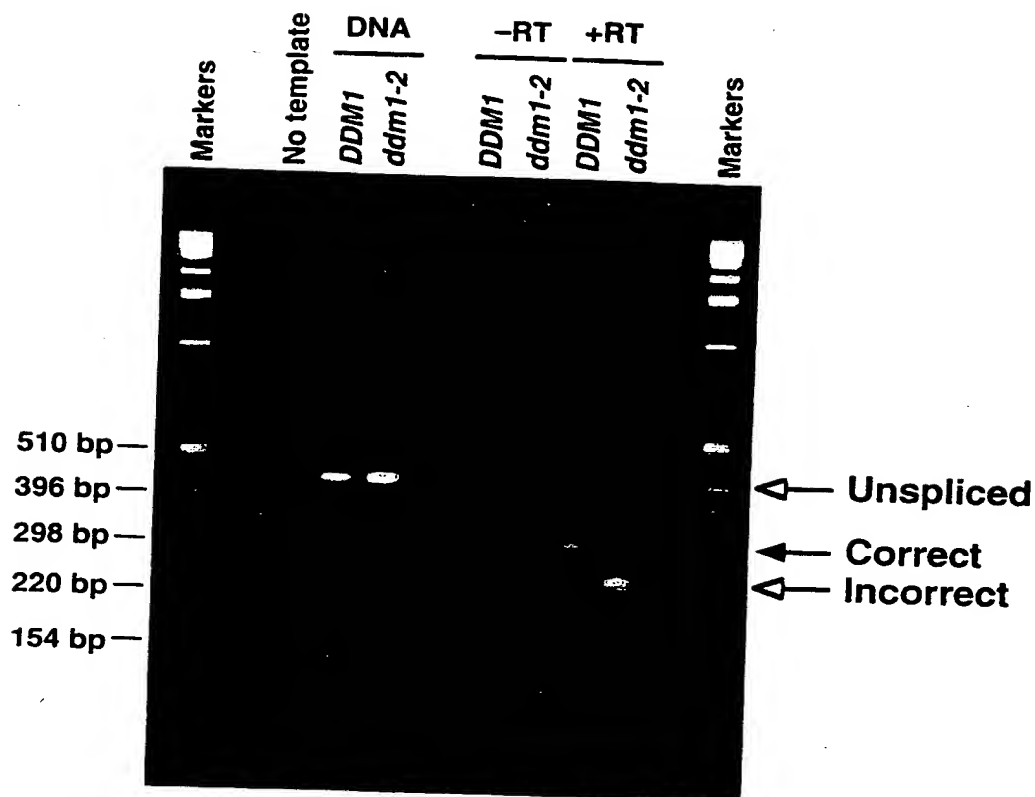


Fig. 2B

SUBSTITUTE SHEET (RULE 26)

[illegible]

Figure 3

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/09268

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/82 C07K14/415 C07K16/16 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JEDDELOH, J.A., ET AL. : "the DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis" GENES AND DEVELOPMENT, vol. 12, no. 11, 1 June 1998 (1998-06-01), pages 1714-1725, XP002114097 the whole document	23,24
X	MITTELSTEN-SCHEID, O., ET AL. : "release of epigenetic gene silencing by trans-acting mutations in Arabidopsis" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 95, January 1998 (1998-01), pages 632-637, XP002114098 cited in the application the whole document	23,24

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

Date of the actual completion of the international search

2 September 1999

Date of mailing of the international search report

15/09/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Holtorf, S

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/09268

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KAKUTANI, T., ET AL. : "developmental abnormalities and epimutations associated with DNA hypomethylation mutations" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 93, October 1996 (1996-10), pages 12406-12411, XP002114099 page 12407, left column; page 12409, left column; Fig. 3 ---	1-6, 10
Y	KAKUTANI, T., ET AL.: "characterization of an Arabidopsis thaliana hypomethylation mutant" NUCLEIC ACID RESEARCH, vol. 23, no. 1, 1995, pages 130-137, XP002049118 cited in the application abstract, last paragraph ---	1-6, 10
A	KAKUTANI, T.: "genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana" THE PLANT JOURNAL, vol. 12, no. 6, 1997, pages 1447-1451, XP002114100 abstract, page 1448, right column ---	1-24
A	ROUNSLEY, S.D., ET AL. : "a BAC end sequence database for identifying minimal overlaps in Arabidopsis genomic sequencing . Update 4." EMBL SEQUENCE DATA LIBRARY, 29 May 1998 (1998-05-29), XP002114101 heidelberg, germany accession no. AQ010627 ---	1-24
A	VONGS, A., ET AL. : "Arabidopsis thaliana DNA-methylation mutants" SCIENCE, vol. 260, June 1993 (1993-06), pages 1926-1928, XP002049119 cited in the application the whole document ---	1-24
A	WO 98 04725 A (UNIV YALE) 5 February 1998 (1998-02-05) abstract, page 10-14; examples 2 + 3, claims; ---	1-24

	-/--	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/09268

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PASZKOWSKI, J., ET AL.: "plant genes: the genetics of epigenetics" CURRENT BIOLOGY, vol. 8, no. 6, March 1998 (1998-03), pages R206-R208, XP002114102 the whole document ---	1-24
P,X	NAKAMURA, Y.: "structural analysis of Arabidopsis thaliana chromosome 5. IX. - unpublished" EMBL SEQUENCE DATA LIBRARY, 7 October 1998 (1998-10-07), XP002114103 heidelberg, germany accession no. AB018119 -----	1,2,10, 15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/09268

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9804725 A	05-02-1998	AU 4048097 A EP 0935666 A	20-02-1998 18-08-1999

Form PCT/ISA/210 (patent family annex) (July 1992)

THIS PAGE BLANK (USPTO)